GRÉGOIRE NICOLIS
ILYA PRIGOGINE

LA COMPLESSITÀ

ESPLORAZIONI NEI NUOVI CAMPI DELLA SCIENZA

GIULIO EINAUDI EDITORE

Negli ultimi decenni è avvenuta in ambito scientifico una rivoluzione che va sotto il nome di «complessità». Il paradigma della fisica classica assumeva di poter ridurre le dinamiche del mondo reale a modelli semplificati, giungendo cosí, con sufficienti livelli di approssimazione, a prevederne sviluppi e comportamenti. L'impredicibilità di sistemi fortemente complessi, secondo questa concezione, derivava unicamente dall'insufficienza delle nostre conoscenze e dei nostri mezzi per dominare tutte le variabili in gioco.

La scienza contemporanea, dalla fisica quantistica allo studio dei sistemi non-lineari, ha prodotto un mutamento di questo punto di vista: i limiti della predicibilità appaiono sempre piú come una caratteristica intrinseca di un gran numero di sistemi del mondo reale. Prendiamo ad esempio un biliardo in cui il movimento iniziale di una biglia influisce sulla posizione di tutte le altre dopo una serie di collisioni: ogni infinitesima variazione iniziale determinerà configurazioni finali molto diverse tra loro, e praticamente impossibili da predire. Immaginando un sistema infinitamente piú grande, ma avente le stesse caratteristiche, come l'atmosfera terrestre, è facile comprendere come, secondo una battuta corrente, il battito d'ala di una farfalla americana possa, concettualmente, provocare un tifone in Cina. Simili sistemi, per quanto intrinsecamente impredicibili, non sono per questo impossibili da modellizzare. Sarà necessario abbandonare i concetti deterministici e prospettare invece modelli «caotici», in grado di descrivere, a un livello superiore, fenomeni apparentemente dominati dal caso.

Per affrontare questi problemi sono necessari nuovi strumenti concettuali: Prigogine e Nicolis li discutono in dettaglio, soffermandosi in particolare sui modelli matematici delle dinamiche caotiche e sui problemi teorici fondamentali che ne derivano, come quelli connessi all'irreversibilità temporale. Su queste premesse è possibile tentare di delineare una teoria unificata della complessità, in grado di ricondurre a un'unica prospettiva sistemi estremamente differenti studiati da discipline anche molto lontane tra loro, come le reazioni dei materiali in condizioni lontane dall'equilibrio, i fenomeni geo-climatici, i comportamenti collettivi delle colonie d'insetti, o quelli degli attori sociali in un sistema economico complesso.

Il libro di Prigogine e Nicolis si presenta come una ricognizione dell'attuale stato dei lavori nello studio della complessità, e insieme come un'introduzione generale a questo complesso di idee che ha ormai cambiato il modo di fare ricerca scientifica.

Grégoire Nicolis insegna fisica chimica teorica all'Università di Bruxelles. I suoi interessi di ricerca vertono principalmente sulla teoria dei processi irreversibili, dei fenomeni non-lineari e dei sistemi complessi.

Ilya Prigogine è direttore degli istituti Solvay di fisica e chimica di Bruxelles e del Centro per lo studio della meccanica statistica e della termodinamica dell'Università di Austin, nel Texas. Nel 1977 ha ricevuto il premio Nobel per i suoi studi sulla teoria delle strutture dissipative. Di Prigogine Einaudi ha pubblicato *Dall'essere al divenire. Tempo e complessità nelle scienze fisiche* (1986) e, con I. Stengers, *La nuova alleanza. Metamorfosi della scienza* (1981).

Grégoire Nicolis e Ilya Prigogine

La complessità

Esplorazioni nei nuovi campi della scienza

Giulio Einaudi editore

Indice

Appendici

Prefazione

Il nostro universo fisico non ha piú come simbolo il moto regolare e periodico dei pianeti, moto che è alla base stessa della meccanica classica. È invece un universo di instabilità e fluttuazioni, che sono all'origine dell'incredibile varietà e ricchezza di forme e strutture che vediamo nel mondo intorno a noi. Abbiamo quindi bisogno di nuovi concetti e nuovi strumenti per descrivere una natura in cui evoluzione e pluralismo sono divenute le parole fondamentali. Scopo di questo libro è offrire una sintetica introduzione ai metodi messi a punto negli ultimi decenni per indagare la *complessità* a ogni livello, dalle molecole ai sistemi biologici, fino alle organizzazioni sociali.

Due discipline in particolare ci sembrano aver svolto un ruolo essenziale nel drammatico mutamento della nostra visione della complessità: la prima è la fisica del non-equilibrio, i cui risultati piú inattesi sono venuti dalla scoperta di nuove fondamentali proprietà della materia in condizioni lontane dall'equilibrio; la seconda è la moderna teoria dei sistemi dinamici, la cui scoperta centrale è la prevalenza dell'instabilità, ossia del fatto che piccoli cambiamenti nelle condizioni iniziali possono determinare grandi amplificazioni degli effetti finali.

Questi nuovi metodi portano a una migliore comprensione dell'ambiente in cui viviamo. Nel mondo che ci circonda troviamo sia inaspettate regolarità, sia fluttuazioni, ugualmente inaspettate, di scala molto ampia. Il fatto che la materia sia associata a una assoluta predominanza delle particelle sulle antiparticelle è un esempio di regolarità, come lo è il fatto che la vita è associata alla predominanza di biomolecole chirali sui loro antipodi. Quale meccanismo selettivo può essere stato all'origine di tali regolarità su larga scala? Viceversa dovremmo aspettarci uniformità e stabilità nelle nostre condizioni climatiche, mentre di fatto, e contrariamente a queste aspettative, il clima ha avuto violente fluttuazioni lungo periodi relativamente brevi se paragonati al tempo di evoluzione del Sole. Come è possibile questo? Cominciamo ora ad avere dei metodi coi quali affrontare questo tipo di domande.

Il capitolo i è dedicato alla presentazione di esempi selezionati di fenomeni complessi sia nel campo della fisico-chimica e dei sistemi biologici, sia piú in generale del mondo che ci circonda. Questa descrizione illustra un certo numero di concetti correlati a meccanismi che si trovano ripetuti in fenomeni differenti, come non-equilibrio, stabilità, biforcazione e rottura di simmetria, ordine a lungo raggio. Nel capitolo ii questi concetti vengono ripresi e analizzati in dettaglio. Essi costituiscono gli elementi basilari di quello che riteniamo essere un nuovo vocabolario scientifico: il *vocabolario della complessità*.

Dopo questi capitoli puramente descrittivi, introdurremo nel capitolo iii il problema della complessità dal punto di vista della moderna teoria dei sistemi dinamici. Discuteremo qui alcuni meccanismi per i quali i sistemi non-lineari allontanati dalla posizione di equilibrio possano generare instabilità che portano a biforcazioni e a rotture di simmetria. Nella nostra analisi un particolare rilievo è dato all'emergere di una *dinamica caotica*, la naturale tendenza di grandi classi di sistemi a evolvere verso stati che mostrano sia le caratteristiche di un comportamento deterministico sia quelle dell'impredicibilità.

Il capitolo iv tenta una descrizione dettagliata dei fenomeni complessi, a un livello piú raffinato di quello puramente fenomenologico dei tre capitoli precedenti. Presenteremo qui gli elementi basilari dell'analisi probabilistica dei sistemi non-lineari e di non-equilibrio, e costruiremo un modello microscopico di biforcazione e di rottura. Discuteremo anche alcuni modi in cui il concetto di informazione può essere integrato nella discussione dei sistemi dinamici.

Nella visione classica vi era una netta distinzione fra caso e necessità, fra comportamento deterministico e comportamento stocastico. L'analisi dei capitoli iii e iv mostra che la questione è molto piú sottile. Ci sono varie forme di casualità, alcune delle quali sono associate al comportamento caotico delle soluzioni di semplici equazioni deterministiche. Nel capitolo v ci porremo la domanda dell'origine della casualità e dell'irreversibilità. Direttamente collegato a questa è il problema della comprensione dell'entropia e, di fatto, lo stesso concetto di tempo. Crediamo di iniziare a decifrare il messaggio della notissima seconda legge della termodinamica: viviamo in un mondo di processi instabili, e questo ci permette di definire una funzione entropia; per di piú viviamo in un mondo in cui la simmetria fra passato e presente si è rotta, un mondo in cui i processi irreversibili portano verso l'equilibrio nel nostro futuro. Questa universale prevalenza della rottura della simmetria temporale è il cuore della seconda legge.

Abbiamo espresso la convinzione che la scienza è destinata a giocare un ruolo sempre piú importante nel nostro sforzo di comprendere il mondo che ci circonda. La capacità di rompere le barriere disciplinari e la ricerca di nuovi modi di considerare problemi anche annosi è dunque uno degli scopi essenziali del metodo di analisi dei fenomeni complessi proposto in questo libro. Il capitolo VI mostrerà come si possa immaginare questo passaggio di conoscenze da un campo all'altro. Dedicheremo quindi una larga parte di questo capitolo a questioni che vanno oltre l'orizzonte dei problemi tradizionali delle scienze fisiche, come i cambiamenti del clima, il comportamento degli insetti sociali e le popolazioni umane. Ognuno di questi problemi ha ovviamente la sua specificità e non si possono assolutamente prospettare ampie generalizzazioni; appare tuttavia evidente che in essi la non-linearità e le fluttuazioni vi giochino un ruolo non secondario. Tutto ciò induce fortemente a pensare che la costruzione dei modelli di questi sistemi dovrebbe trarre beneficio delle nuove prospettive che lo studio dei fenomeni complessi nei sistemi dinamici non-lineari ha portato alla scienza.

Le appendici, infine, sono dedicate a un'analisi piú strettamente quantitativa di alcune delle tecniche utilizzate nei capitoli principali.

Durante la preparazione di questo libro abbiamo beneficiato grandemente delle discussioni avute con numerosi colleghi e collaboratori. Siamo particolarmente grati a E. Rebhan per aver letto criticamente il manoscritto originale, suggerendo numerosi miglioramenti, e per averne coordinato il lavoro di traduzione in tedesco. Con altrettanto piacere vogliamo qui ricordare i suggerimenti e l'aiuto datoci da C. Baesens, F. Baras, J. L. Deneubourg, Y. Elskens, R. Feistel, H. Frisch, R. Mazo, M. Malek Mansour, C. Nicolis, J. S. Nicolis, G. Rao e S. Rao.

Infine ringraziamo P. Pape, A. Kimmons, S. Wellens e M. Adam per il difficile compito di battitura di due successive versioni del manoscritto, e ancora P. Kinet per la sua efficiente assistenza tecnica.

La ricerca degli autori nel campo dei fenomeni irreversibili e della dinamica non-lineare beneficia del sostegno degli International Solvay Institutes of Physics and Chemistry, dei ministeri belgi dell'Educazione e della Ricerca scientifica, dalla Commissione delle Comunità Europee, dalla fondazione R. A. Welch di Houston (Texas), e dal dipartimento dell'Energia del governo degli Stati Uniti.

G. NICOLIS, I. PRIGOGINE

Prologo

La scienza in un'era di transizione

Quali che siano le nostre preoccupazioni professionali, non possiamo non avere la sensazione di vivere in un'era di transizione. Dobbiamo trovare ed esplorare nuove risorse; capire meglio il nostro ambiente; raggiungere una coesistenza meno distruttiva con la natura. L'intervallo di tempo per le modifiche qualitative che richiedono questi obiettivi principali non è paragonabile all'immenso arco di tempo necessario all'evoluzione biologica e geologica. Esso è dell'ordine del decennio, e pertanto interferisce con la vita di ognuno di noi e con le vite di coloro che appartengono alla prossima generazione.

Non è possibile anticipare che cosa nascerà da questo periodo di transizione, ma è chiaro che la scienza è obbligata a giocare un ruolo sempre piú importante nello sforzo di affrontare la sfida di capire e di dare nuova forma all'ambiente che ci circonda. Colpisce il fatto che in questo momento cruciale la scienza stessa stia attraversando un periodo di riconcettualizzazione.

Le due grandi rivoluzioni in fisica all'inizio di questo secolo furono la meccanica quantistica e la relatività. Entrambe partirono come correzioni della meccanica classica, rese necessarie dalla scoperta del ruolo delle costanti universali c (la velocità della luce) e h (la costante di Planck). Oggi entrambe hanno assunto un inaspettato carattere «temporale»: la meccanica quantistica tratta, nelle sue parti piú interessanti, la descrizione delle particelle instabili e le loro mutue trasformazioni. Analogamente, la relatività, che iniziò come una teoria geometrica, oggi è una teoria principalmente associata alla storia termica dell'universo.

Le particelle elementari e la cosmologia corrispondono a condizioni estreme: esse fanno parte della fisica delle alte energie. Tuttavia anche nel mondo macroscopico, la fisica sta subendo radicali trasformazioni. Fino a pochi anni fa, un fisico, se richiesto di dire che cosa è noto e che cosa non lo è, avrebbe risposto che ci sono veri problemi solo alle fron-

tiere del nostro universo, a livello cioè delle particelle elementari e a livello della cosmologia; inversamente, avrebbe affermato che le leggi fondamentali che sono rilevanti a livello macroscopico sono ben note. Oggi una crescente minoranza mette in dubbio questa visione ottimistica. Anche a questo livello si vede che alcune delle domande fondamentali rimangono ampiamente senza risposta.

La storia della scienza nei tre secoli che seguirono la sintesi newtoniana è in effetti una storia drammatica. Vi sono stati momenti in cui il programma della scienza classica sembrava vicino al completamento: sembrava infatti avvicinarsi a un livello fondamentale, che era il veicolo delle leggi deterministiche e reversibili. Ognuna di queste volte, tuttavia, qualcosa non andò per il verso giusto. Il progetto doveva essere allargato, e il livello fondamentale diveniva impreciso.

Oggi, ovunque guardiamo troviamo evoluzione, diversificazione e instabilità. Sappiamo da lungo tempo di vivere in un mondo pluralistico, nel quale troviamo fenomeni sia deterministici che stocastici, sia reversibili che irreversibili. Osserviamo i fenomeni deterministici come il pendolo senza attrito o la traiettoria della Luna attorno alla Terra; per di piú sappiamo che il pendolo senza attrito è reversibile, poiché futuro e passato giocano lo stesso ruolo nelle equazioni che descrivono il moto. Ma vi sono anche processi irreversibili, come la diffusione o le reazioni chimiche; qui vi è una direzione privilegiata del tempo: il sistema diviene uniforme col passare del tempo. In piú, se vogliamo evitare il paradosso di attribuire la varietà dei fenomeni naturali a un progetto impresso già al momento del Big Bang, siamo forzati ad ammettere l'esistenza di processi stocastici.

Ciò che è cambiato dall'inizio di questo secolo è la nostra valutazione della relativa importanza di questi quattro tipi di fenomeni: reversibile contro irreversibile, deterministico contro stocastico.

All'inizio di questo secolo continuando la tradizione della ricerca secondo il programma classico, i fisici erano quasi unanimi nell'ammettere che le leggi fondamentali dell'universo erano deterministiche e reversibili. I processi che non si adattavano a questo schema erano pensati come delle eccezioni, semplici artefatti dovuti alla complessità, che doveva essa stessa essere spiegata invocando la nostra ignoranza o la nostra mancanza di controllo delle variabili coinvolte. Ora che il secolo volge alla fine, siamo sempre piú numerosi nel pensare che molti processi fondamentali della natura sono irreversibili e stocastici e che le leggi deterministiche e reversibili che descrivono le interazioni elementari non dicono l'intera verità. Questo fatto porta a una nuova visione

delle cose: non piú passiva, come descritta nella visione meccanicistica del mondo, ma associata a una spontanea attività. Questa nuova visione è cosí profonda che pensiamo di poter parlare di un nuovo dialogo dell'uomo con la natura.

È interessante indagare su come un simile cambiamento sia potuto avvenire in un cosí breve lasso di tempo. È la conseguenza di risultati inattesi, ottenuti in campi molto diversi della fisica e della chimica come le particelle elementari, la cosmologia, e lo studio delle autoorganizzazioni nei sistemi lontani dall'equilibrio.

Chi avrebbe creduto, cinquant'anni fa, che la maggior parte delle particelle elementari, e forse tutte, sono instabili? O che possiamo parlare dell'evoluzione dell'universo come un tutt'uno? O che, lontano dall'equilibrio le molecole possono comunicare (usando un termine antropomorfico), cosí come è testimoniato dagli orologi chimici?

Tutte queste inattese scoperte ebbero anche un drastico effetto sul nostro modo di vedere le relazioni fra scienza «rigida» e scienza «morbida». Secondo la visione classica vi è una netta distinzione fra sistemi semplici, come quelli studiati in fisica e chimica, e i sistemi complessi, come quelli studiati in biologia e nelle scienze umane. Invero, non si potrebbe immaginare un piú grande contrasto di quello esistente fra i semplici modelli di dinamica classica o il semplice comportamento di un gas o un liquido, e i complessi processi che scopriamo nell'evoluzione della vita o nella storia delle società umane.

Proprio perché questo divario si sta assottigliando, possiamo ora pensare di trasferire delle conoscenze a situazioni alle quali non possono venire applicati i concetti della fisica classica.

Nella fisica classica, l'osservatore era fuori dal sistema. Era colui che decideva, mentre il sistema era sottomesso a leggi deterministiche. In altri termini, si aveva da un lato qualcuno che «decideva», che era «libero», e dall'altro i componenti di un sistema, fossero essi individui o organizzazioni, che dovevano rispettare qualunque piano del capo.

Stiamo allontanandoci sempre piú da questa dicotomia. Non solo nelle scienze umane ma anche in fisica sappiamo che siamo sia spettatori che attori, per usare una ben nota espressione di Niels Bohr. Al posto di una costruzione nella quale il presente implica il futuro, andiamo verso un mondo nel quale il futuro è aperto; nel quale il tempo è una costruzione dentro cui tutti possiamo precipitare.

Capitolo primo

La complessità nella natura

Qual è la differenza, se esiste, fra il pendolo e il battito del cuore o fra un cristallo d'acqua e un fiocco di neve? È il mondo dei fenomeni fisici e chimici, un mondo semplice e prevedibile, dove tutti i fatti osservati possono essere adeguatamente interpretati appellandosi a poche fondamentali interazioni? E, alternativamente, la complessità è confinata esclusivamente alla biologia? Queste e altre domande verranno sollevate lungo il presente capitolo introduttivo. L'esame dei dati sperimentali ci porteranno alla conclusione che la distinzione fra fenomeni fisico-chimici e fenomeni biologici, fra comportamento «semplice» e comportamento «complesso», non è poi cosí sottile come si potrebbe pensare intuitivamente. Questa osservazione ci suggerirà una *visione pluralistica* del mondo fisico, dove diversi tipi di fenomeni coesistono fianco a fianco quando vengono variate le condizioni cui è soggetto il sistema. La visione di un mondo aperto è il messaggio principale di questo libro.

1. *Che cos'è la complessità?*

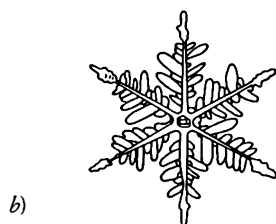
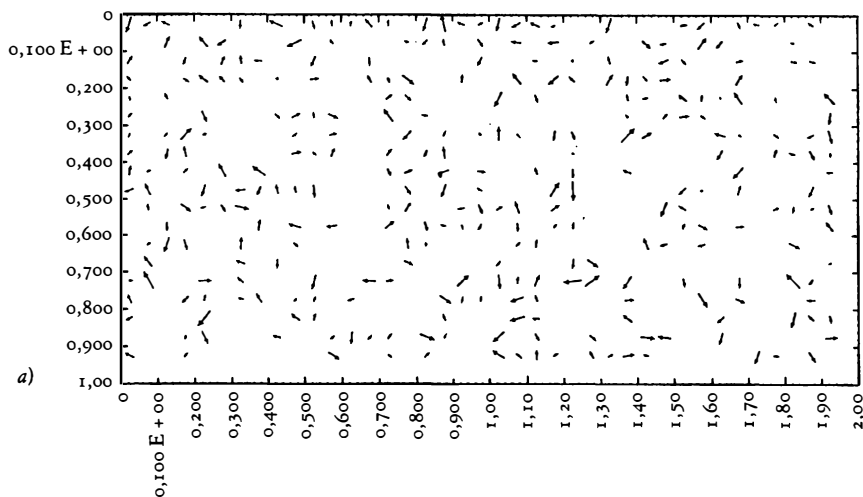
La complessità è una nozione che appartiene al nostro vocabolario quotidiano. Noi l'abbiamo usata fin dalla nostra infanzia in diversi contesti, sebbene avessimo l'intuizione che la complessità fosse qualcosa di correlato con le varie manifestazioni della vita. Cosí, per la maggior parte di noi, i fenomeni tradizionalmente trattati nei testi di fisica, come la caduta libera di un grave, o il moto del pendolo, sono fondamentalmente «semplici». Al contrario, saremmo tutti pronti nell'asserire che il nostro sistema economico, il nostro linguaggio, il cervello dei mammiferi, o anche il piú umile dei batteri siano «sistemi complessi».

Ma cosa possiamo dire di un centimetro cubo di un gas o di un liquido, per esempio di acqua? Come nei precedenti esempi, siamo anche qui in presenza di sistemi contenenti un grande numero di elementi in-

teragenti fra loro, le molecole. È difficile immaginare l'entità di questo numero. Dalla fisica chimica elementare sappiamo che alla pressione di 760 mm di mercurio e alla temperatura di 0 °C, una mole di un qualsiasi gas contiene un numero di molecole uguale al numero di Avogadro,

Figura 1.1

a) Visualizzazione del caos molecolare. La figura sintetizza i risultati di una simulazione al calcolatore delle equazioni di moto di 400 dischi rigidi elastici in una scatola rettangolare delle dimensioni di 120 × 60 (diametri molecolari)[2]. Le condizioni al contorno sono la temperatura sui due estremi orizzontali, fissata allo stesso valore ($T = 1$ in unità in cui $k_B/m = 1/2$, dove k_B è la costante di Boltzmann, e m è la massa molecolare), e condizioni periodiche sugli estremi verticali. Il sistema viene integrato per circa 500 collisioni per particella, ed ha raggiunto l'equilibrio termico per ogni scopo pratico. Le frecce iniziano nella posizione centrale di ciascuna particella. La loro direzione e lunghezza indicano, rispettivamente, la direzione e magnitudine della velocità della corrispondente particella. (Per gentile concessione di F. Baras).

b) Un cristallo di neve che appare nella sua caratteristica forma dendritica.

6,025 o 10^{23} molecole. Questo ci dà 10^{19} molecole, 10 miliardi di miliardi, ammassate in uno spazio di un centimetro cubo, che si muovono in tutte le possibili direzioni, e che collidono continuamente fra loro.

È abbastanza questo per dire che questo sistema è «complesso»? Nonostante gli impressionanti numeri sopra menzionati, la nostra intuizione ci dice che la risposta a questa domanda dovrebbe essere negativa a causa dell'assenza di una qualsiasi attività coordinata, forma o dinamica. Siamo infatti tentati a guardare un simile sistema come il prototipo di un comportamento disordinato, irregolare, frequentemente chiamato dai fisici come *caos molecolare*, in cui le molecole si muovono incoerentemente perché non si possono riconoscere l'una con l'altra a distanze superiori a pochi angstrom (Å; 1 Å = 10^{-8} un centimilionesimo di centimetro). Si deve attribuire questo fatto a una fondamentale proprietà delle forze intermolecolari delle interazioni in questo e in molti altri sistemi che si incontrano in natura, ovvero il *carattere a corto raggio* di queste forze.

Supponiamo ora che questo centimetro cubo di fluido sia soggetto alle ordinarie condizioni di una tempesta invernale. Si otterrà un magnifico fiocco di neve (figura 1.1*b*) con la sua caratteristica forma dendritica. Ora potremmo essere pronti a parlare di complessità, contemplando l'inimitabile lavoro che ha realizzato la natura. Possiamo cosí vedere che lo stesso sistema può apparire sotto forme diverse, che evocano impressioni di «semplicità» e successivamente di «complessità».

Questo esempio ci insegna un'importante lezione: è piú naturale, o meno ambiguo, parlare di *comportamento complesso* piuttosto che di sistemi complessi. Lo studio di un simile comportamento rivelerà un certo numero di caratteristiche comuni alle diverse classi di sistemi e ci permetterà di arrivare a una giusta comprensione della complessità. Questo compito costituisce lo scopo principale del nostro saggio.

2. *Autoorganizzazione nei sistemi fisico-chimici: la nascita della complessità.*

La scienza è soprattutto un'attività sperimentale, perciò dobbiamo guardarci attorno per avere un'idea della generalità e dell'importanza dei fenomeni complessi.

Abbiamo già sottolineato come il fascino esercitato dalla biologia su ognuno di noi sia responsabile di una diffusa identificazione della complessità con la vita. Sorprendentemente, proprio questa idea perenne

sarà la prima a cadere nella nostra strada verso la comprensione della complessità. In effetti, fin dagli anni '60 siamo testimoni di una rivoluzione sia nelle scienze matematiche che nelle scienze fisiche, che impone un *nuovo atteggiamento nella descrizione della natura*. Gli sviluppi paralleli nella teoria termodinamica dei fenomeni irreversibili, nella teoria dei sistemi dinamici e nella meccanica classica, sono alla fine confluiti per mostrare in modo inoppugnabile che il divario fra «semplice» e «complesso», fra «ordine» e «disordine», è molto piú sottile di quanto prima non si supponesse. Oggi si sa che già i piú semplici esempi di sistemi meccanici presentano un comportamento complesso. Un pendolo forzato – l'altalena – nel confine fra vibrazione e rotazione dà luogo a una ricca varietà di moti, inclusa la possibilità di escursioni casuali di tipo turbolento dalla sua posizione di equilibrio. Ordinari sistemi come uno strato di fluido o una miscela di prodotti chimici possono generare, sotto certe condizioni, *fenomeni autoorganizzati* in scala macroscopica sotto forma di modelli spaziali di ritmi temporali. In poche parole la complessità non è piú limitata alla biologia; sta invadendo la scienza fisica e sembra profondamente radicata nelle leggi della natura.

Come risultato di queste scoperte, l'interesse verso la fisica macroscopica, la fisica che tratta i fenomeni che ci circondano, sta crescendo enormemente. La descrizione di particolari fenomeni rappresentativi che serviranno come riferimento continuo lungo tutto il volume, verrà presa in esame nei prossimi paragrafi di questo capitolo.

3. *Convezione termica, un prototipo dell'autoorganizzazione in fisica.*

Tratteremo dapprima i movimenti macroscopici dei fluidi (liquidi o gas) sotto l'effetto delle disomogeneità termiche, conosciuti come *convezione termica*.

Lo studio di questi moti, anche oltre la parte specifica che ci riguarda, è lontano dall'essere accademico. Senza nemmeno contare le numerose applicazioni tecnologiche, la convezione termica è alla base di numerosi fenomeni importanti e spettacolari nel nostro pianeta. Esempi sono la circolazione dell'atmosfera e degli oceani (che determinano in larga misura i cambiamenti climatici a corto e a medio termine) e la deriva dei continenti (il moto delle placche continentali è indotto dai moti in larga scala del mantello). Poco piú lontano da noi, la convezione è alla base del trasferimento di calore e della materia nel Sole, e dunque incide fortemente sulla sua attività. In laboratorio, possiamo studiare il

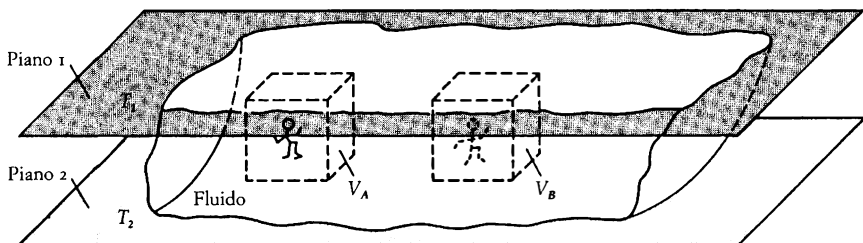
meccanismo della convezione termica su un apparato sperimentale di dimensioni molto piú modeste. Il seguente semplice esperimento, realizzato per la prima volta da Bénard, ci fa osservare un certo numero di proprietà sorprendenti.

Immaginiamo uno strato di fluido (per esempio l'acqua) limitato da due piani orizzontali e paralleli le cui dimensioni laterali siano molto piú lunghe della larghezza dello strato. Lasciato a se stesso, il fluido tenderà rapidamente verso uno stato omogeneo nel quale, parlando statisticamente, tutte le sue parti sono identiche. Per esempio, un minuscolo osservatore non sarà in grado di sapere se egli si trova nel piccolo volume V_A piuttosto che nel piccolo volume V_B del fluido sulla sola base delle osservazioni dell'ambiente che lo circonda (figura 1.2). Tutti i volumi che si possono definire arbitrariamente entro un fluido sono indistinguibili, e la conoscenza dello stato di uno di questi è sufficiente per conoscere lo stato di tutti, indipendentemente dalla loro forma e dimensione. In altre parole, dal punto di vista del nostro osservatore, la posizione che egli occupa non fa differenza. Alternativamente, non c'è alcun modo intrinseco che lo renda capace di percepire la nozione di spazio.

L'omogeneità di questo sistema si estende naturalmente a tutte le proprietà e, in particolare, alla sua temperatura, che sarà la stessa in tutte le parti del fluido e uguale alla temperatura dei piani che lo racchiudono, o, alternativamente, alla temperatura del «mondo esterno».

Tutte queste proprietà sono caratteristiche di un sistema in un particolare stato, lo stato per cui non c'è né un movimento massiccio, né una differenza di temperatura con il mondo esterno. Possiamo esprime-

Figura 1.2

Un minuscolo osservatore che guarda lo stato degli elementi di volume V_A, V_B entro un recipiente di liquido in equilibrio li trova indistinguibili e conclude che il fluido mostra una invarianza traslazionale lungo la direzione orizzontale.

re questa proprietà in maniera quantitativa come segue. Chiamiamo T_1 e T_2 le temperature dei piani 1 e 2 rispettivamente; la differenza di temperatura, ΔT, è definita $\Delta T = T_2 - T_1 = 0$, e poiché all'equilibrio non c'è differenza di temperatura dobbiamo necessariamente avere

$$\Delta T_e = T_2 - T_1 = 0. \tag{1.1}$$

Immaginiamo ora che qualcuno ponga il dito su un piano per un breve istante. La temperatura in quella parte del piano verrà momentaneamente modificata (per esempio, da 20°C alla temperatura del corpo umano 36,5°C). Un episodio come questo, che avviene per caso in un sistema e modifica localmente (e in generale debolmente) qualcuna delle sue proprietà, è detto *perturbazione*. Per il nostro sistema all'equilibrio questa perturbazione termica non avrà influenza, poiché la temperatura diverrà nuovamente uniforme e uguale al suo valore iniziale (20°C nel nostro esempio). In altre parole la perturbazione scompare; nel sistema non c'è traccia di essa. Quando un sistema si trova in uno stato in cui le perturbazioni che agiscono su di esso scompaiono in un tempo piú o meno breve, diciamo che lo stato è *asintoticamente stabile*.

Dal punto di vista del nostro minuscolo osservatore, non solo l'omogeneità del fluido gli rende impossibile sviluppare un intrinseco concetto di spazio, ma, in aggiunta, la stabilità dello stato di equilibrio rende alla fine tutti gli istanti identici. È dunque impossibile per lui sviluppare un intrinseco concetto di tempo. Non è possibile parlare di «comportamento» per un sistema in una simile *semplice* situazione cosicché allontaniamoci da essa.

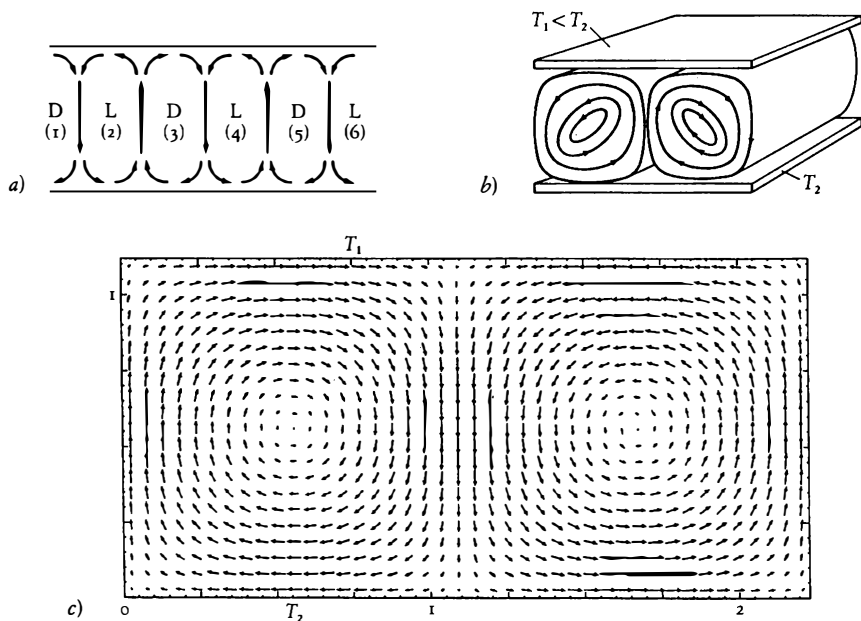
Questo si può realizzare per esempio scaldando lo strato di fluido dal basso. Cosí facendo, comunichiamo energia al nostro sistema sotto forma di calore. Per di piú, quando la temperatura T_2 del piatto sottostante sarà piú alta della T_1, la condizione di equilibrio dell'equazione 1.1 sarà violata ($\Delta T > 0$). In altre parole non permettiamo al sistema di raggiungere l'equilibrio, applicando un *vincolo esterno*. Si noti che nel presente esempio il vincolo esterno implica flusso di energia e viceversa.

Supponiamo dapprima che il vincolo sia debole (ΔT piccolo). Il nostro sistema adotterà di nuovo un semplice e unico stato, nel quale l'unico processo che si attua sarà il trasporto di calore dal piatto inferiore a quello superiore, dal quale il calore uscirà nel mondo esterno per assicurare che T_1 rimanga costante. L'unica differenza con lo stato di equilibrio sarà che la temperatura, e attraverso di essa la densità e la pressio-

ne, non saranno piú uniformi. Esse varieranno, in modo praticamente lineare, dalle regioni calde (sotto) alle regioni fredde (sopra). Questo fenomeno è noto come *convezione termica*. In questo nuovo stato che il sistema ha raggiunto per rispondere ai vincoli, la stabilità prevarrà ancora e il comportamento sarà «semplice» come all'equilibrio.

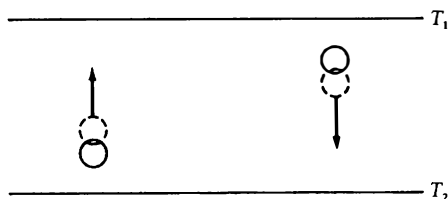
Figura 1.3

a) e *b*) Due visioni delle celle di convezione (celle di Bénard). Si noti l'opposta direzione di rotazione di due celle adiacenti, schematizzata in *c*).

c) Complessità e ordine a lungo raggio a partire dal caos molecolare in un sistema sottoposto a vincoli di non-equilibrio. Le equazioni di moto di 5000 dischi rigidi elastici in una scatola rettangolare che misura circa 112 × 56 (diametri molecolari)² sono state risolte al calcolatore con le seguenti condizioni al contorno: temperature fissate rispettivamente agli estremi orizzontali inferiore e superiore, $T_2 = 1{,}61$, $T_1 = 0{,}51$ (in unità ridotte in cui la velocità del suono nel mezzo è uguale a uno); e pareti verticali perfettamente riflettenti. Il sistema viene dapprima integrato per un intervallo di tempo corrispondente al verificarsi di piú di 30×10^6 collisioni. In seguito le velocità molecolari istantanee vengono registrate per ulteriori 20×10^6 collisioni, e vengono rappresentate graficamente le loro medie su questo intervallo di tempo e su una piccola cella spaziale contenente cinque o sei molecole. Il significato delle frecce è lo stesso che nella figura 1.1*a*. Si noti che, nella situazione di equilibrio della figura 1.1*a*, una simile media darebbe luogo ovunque a un campo di velocità zero. (Da M. Mareschal, M. Malek Mansour, A. Poul e E. Kestemont, *Phys. Rev. Letters*, 61, 2550, 1988).

Rimuovendo il sistema dall'equilibrio sempre di piú, attraverso un aumento del ΔT, osserviamo che improvvisamente, a un valore di ΔT che chiameremo *critico*, ΔT_c, la materia inizia a manifestare un massiccio movimento. Per di piú, questo movimento è lontano dalla casualità: il fluido è strutturato in una serie di «celle» (figura 1.3), conosciute come celle di Bénard. Questo è il regime della convezione termica definito all'inizio del presente capitolo.

La figura 1.4 mostra una spiegazione qualitativa del fenomeno. A causa dell'espansione termica il fluido diventa stratificato, poiché le parti vicine al piano inferiore sono caratterizzate da una densità piú bassa di quelle vicine al piano superiore. Questo fatto dà luogo a un gradiente di densità che si oppone alla forza di gravità. Si può facilmente comprendere come questa configurazione sia potenzialmente instabile: consideriamo un piccolo volume di fluido vicino al piano inferiore. Immaginiamo che questo elemento di volume sia debolmente spostato verso l'alto da una perturbazione. Poiché ora si trova in una regione piú fredda e dunque piú densa subirà una spinta di Archimede verso l'alto che tenderà ad amplificare ulteriormente il movimento ascendente. Se d'altra parte, una piccola goccia inizialmente vicina al piano superiore si sposta verso il basso, penetrerà in una regione di densità inferiore, e la spinta di Archimede tenderà ad amplificare ulteriormente il movimento verso il basso. Vediamo dunque che, in linea di principio, il fluido può generare correnti ascendenti e discendenti come quelle osservate nell'esperimento. La ragione per cui queste correnti non appaiono non appena ΔT non è zero, come si potrebbe pensare da quanto detto piú sopra, è da ricercarsi nel fatto che gli effetti destabilizzanti sono

Figura 1.4

Spiegazione qualitativa sull'origine della convezione termica. Se una porzione del fluido caldo vicina al fondo si sposta leggermente verso l'alto, entra in una regione di densità media maggiore e subisce pertanto una forza ascensionale. Analogamente una porzione di fluido freddo vicina al bordo superiore se spostata leggermente verso il basso diviene piú pesante delle circostanti e tende ad affondare.

controbilanciati dagli effetti stabilizzanti della viscosità del fluido, che genera un attrito interno che si oppone al movimento, cosí come dalla conduzione termica, che tende a livellare la differenza di temperatura fra la goccia spostata e il suo ambiente circostante. Questo spiega l'esistenza di una soglia critica, ΔT_c, osservabile nell'esperimento. La figura 1.3*b* mostra ancora di piú la *complessità* dei movimenti: a un certo punto il fluido sale, si muove lungo il piano 1, poi scende verso il basso, si muove lungo il piano 2, e quindi sale nuovamente... Le celle si allargano, lungo l'asse orizzontale, adottando successivamente una rotazione destrogira (D) o levogira (L).

Si potrebbe obiettare che questa è una complessità modesta in confronto con quella di un semplice batterio. Ma torniamo ora al nostro minuscolo osservatore. Al suo livello, il suo universo è stato completamente trasformato. Per esempio egli può stabilire dove si trova e dove non si trova osservando il senso di rotazione della cella che occupa. Per di piú, contando durante il suo spostamento, il numero di celle che attraversa, può acquisire un'efficiente nozione di spazio. Chiamiamo *rottura di simmetria* questa apparizione della nozione di spazio in un sistema nel quale, fino a un certo momento, questa nozione non poteva essere percepita in modo intrinseco. In un certo senso, la rottura di simmetria ci porta da una visione dello spazio statica, geometrica, a una visione «aristotelica», in cui lo spazio è modellato dalle funzioni che si stanno svolgendo nel sistema.

Forse l'aspetto piú notevole da rilevare in questa improvvisa transizione da un comportamento semplice a uno complesso è l'*ordine* e la *coerenza* del sistema. Quando ΔT era al di sotto del valore critico ΔT_c, l'omogeneità del fluido nella direzione orizzontale rendeva le diverse parti indipendenti le une dalle altre. Era assolutamente indifferente scambiare il volume V_A con V_B (figura 1.2). Nel volume V_C, situato fra V_A e V_B, non si osserva nessuna modifica delle proprietà osservabili per il fatto che V_A è alla sua destra anziché alla sua sinistra. Al contrario, oltre la soglia ΔT_c, tutto avviene come se ogni volumetto stesse guardando il comportamento dei suoi vicini, e ne «tenesse conto» per giocare adeguatamente il suo ruolo e per partecipare al quadro d'insieme. Questo suggerisce l'esistenza di correlazioni, ovvero di relazioni statisticamente riproducibili fra parti distanti del sistema. Arriviamo a questa conclusione chiave osservando il fenomeno. Questo argomento verrà ripreso per intero e trattato in dettaglio nei capitoli II e IV. Tuttavia, già a questo primo stadio della nostra analisi, si dovrebbe ormai confrontare il carattere a lungo raggio di queste correlazioni con il corto raggio delle

forze intermolecolari. La dimensione spaziale caratteristica di una cella di Bénard nelle usuali condizioni di laboratorio è dell'ordine del millimetro (10^{-1} cm), mentre le dimensioni caratteristiche delle forze intermolecolari sono dell'ordine dell'Amstrong (10^{-8} cm). Le forze intermolecolari agiscono fino a una distanza all'incirca pari alle dimensioni di una molecola; una cella di Bénard comprende qualcosa come 10^{20} molecole. Che questo numero enorme di particelle si possa comportare in modo coerente nonostante il moto termico casuale seguito da ciascuna di esse, è una delle principali proprietà che caratterizzano il manifestarsi del comportamento complesso. La figura 1.3 fornisce una visualizzazione dell'emergere di questa complessità «organizzata» dall'interazione tra il moto termico disordinato delle singole molecole e il vincolo del non-equilibrio. Il confronto con la figura 1*a* è illuminante.

Nelle ultime pagine abbiamo usato un vocabolario che include nozioni come coerenza, complessità, ordine, che sono da lungo tempo parte integrante della biologia, ma che, fino a tempi recenti, erano al di fuori dalla corrente principale della fisica. La possibilità di descrivere, attraverso questi concetti fondamentali, il comportamento sia degli esseri viventi che degli ordinari sistemi fisici, è uno dei principali sviluppi che la storia della scienza non avrebbe assolutamente potuto prevedere fino a pochi anni fa.

Ma le celle di Bénard ci riservano altre sorprese. Da una parte, l'esperimento è perfettamente riproducibile, poiché, realizzando le stesse condizioni sperimentali, si vedrà sempre apparire il moto convettivo allo stesso valore di soglia ΔT_c. Ma dall'altra parte, come vediamo dalla figura 1.3, la materia è strutturata in celle che sono alternativamente destrogire e levogire. Una volta stabilita una direzione di rotazione, essa rimarrà identica a se stessa. Tuttavia, per quanto possa essere sofisticato il controllo dell'apparato sperimentale, si possono realizzare due situazioni qualitativamente diverse subito oltre la soglia critica ΔT_c[1]. Nella figura 1.3*a* la cella 1 è destrogira (e perciò la cella 2 è levogira, la 3 destrogira, ecc.), ma potrebbe essere stato altrimenti, e allora la cella 2 risulterebbe destrogira, con i corrispondenti cambiamenti per le altre celle. Non appena ΔT supera leggermente ΔT_c, sappiamo perfettamente che appariranno le celle; questo fenomeno è dunque soggetto a uno

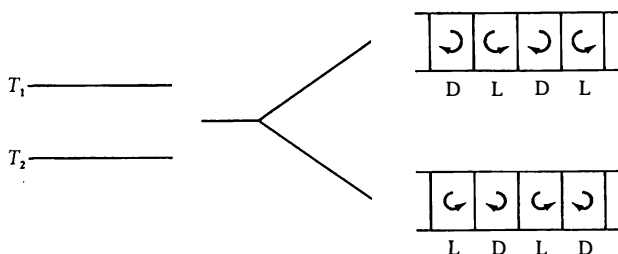
[1] È un dato di fatto che esistono *infinite* possibilità, poiché per un grande sistema l'intera struttura può essere traslata rigidamente lungo la direzione orizzontale (rottura di un *gruppo continuo di simmetria*). Tuttavia la manifestazione caratteristica di questa molteplicità rimane la direzione di rotazione della cella, e per questa ragione usiamo questo criterio durante tutto il paragrafo.

stretto determinismo. Al contrario, la direzione delle celle è impredicibile e incontrollabile: solo il caso, sotto forma della particolare perturbazione che ha prevalso al momento dell'esperimento, può decidere se una certa cella è destrogira o levogira. Arriviamo cosí a una notevole cooperazione fra determinismo e caso, che ci ricorda una simile dualità familiare nella biologia fin dall'era di Darwin (mutazione - selezione naturale), e che è stata fino ad ora limitata, nelle scienze fisiche, alla descrizione quantistica dei fenomeni che avvengono su scala microscopica.

Vediamo che lontano dall'equilibrio, ovvero quando il vincolo è sufficientemente forte, il sistema si adatta a ciò che lo circonda in molti modi diversi, o per essere meno antropomorfici, diciamo che *piú soluzioni* sono possibili per gli *stessi valori parametrici.* Solo il caso deciderà quale di queste soluzioni verrà realizzata. Il fatto che, fra molte scelte solo una venga realizzata conferisce al sistema una *dimensione storica*, una sorta di «memoria» di un evento passato che ha avuto luogo in un momento critico e che inciderà sulla sua ulteriore evoluzione.

Cosa succede quando il vincolo termico aumenta molto al di sopra della prima soglia di strutturazione? Per una certa gamma di valori le celle di Bénard verranno mantenute globalmente, ma qualcuna delle loro caratteristiche specifiche verrà modificata. Poi improvvisamente, oltre un altro valore critico, $\Delta T_c'$, osserviamo una nuova forte manifestazione di casualità: la struttura diventerà confusa e apparirà un regime caratterizzato da una irregolare dipendenza delle variabili dal tempo. Questo costituisce il precursore di ciò che gli ingegneri e i fluido-dinamici chiamano da circa un secolo *turbolenza.* Piú in generale, ora sem-

Figura 1.5

Molteplicità di soluzioni oltre la soglia della instabilità della convezione termica. A seconda delle condizioni, una certa regione dello spazio può far parte della cella destrogira o levogira.

bra che la turbolenza sia un aspetto della tendenza generale di molte classi di sistemi a evolvere in modo *caotico* sotto certe condizioni. Discuteremo ulteriormente questo punto nel capitolo III.

Per riassumere, abbiamo visto che il non-equilibrio ha messo in grado il sistema di evitare il disordine termico descritto nella figura 1.1*a* e di trasformare parte dell'energia comunicata dall'ambiente circostante in un comportamento ordinato di un nuovo tipo, la *struttura dissipativa*: un regime caratterizzato dalla rottura di simmetria, dalle scelte multiple, e da correlazioni a raggio macroscopico. Possiamo dunque dire letteralmente che abbiamo assistito alla nascita della complessità. È vero che il tipo di complessità raggiunto è piuttosto modesto, ma presenta tuttavia caratteristiche che erano esclusivo retaggio dei sistemi biologici. Cosa ancora piú importante, la complessità appare non come una sfida alle leggi della fisica, ma come una inevitabile conseguenza di esse, quando si realizzano le appropriate condizioni.

4. *Autoorganizzazione in chimica.*

Abbiamo trattato le celle di Bénard come parte della fisica, secondo la tradizione, principalmente perché la natura chimica delle sostanze che costituiscono lo strato rimane immutato durante il fenomeno. Dobbiamo ora considerare le reazioni chimiche, ovvero i processi che coinvolgono modificazioni nell'identità delle molecole costituenti. Tali fenomeni sono di interesse generale. Buona parte dell'industria chimica, è basata sulla catalisi eterogenea, attraverso la quale alcuni passi necessari per la sintesi di un prodotto sono accelerati dalla presenza di una superficie reagente con il mezzo. Per esempio, l'ossidazione dell'ammoniaca avviene normalmente in presenza di un catalizzatore al platino, mentre nella decomposizione del N_2O si utilizza una superficie catalitica di ossido di rame. La combustione, come per esempio quella degli idrocarburi, grazie alla quale funzionano i motori termici, è un'altra classe importante di trasformazioni chimiche. In aggiunta, molte manifestazioni di ordine biologico si possono ricondurre a reazioni chimiche nelle quali sono presenti speciali catalizzatori, gli enzimi.

In una tipica reazione chimica una molecola di specie A (per esempio la molecola idrossido OH) si può combinare con una molecola di specie B (per esempio la molecola idrogeno H_2) per produrre una molecola di specie C e una molecola di specie D (nel nostro esempio rispetti-

vamente H_2O e idrogeno atomico H). Questo processo in simboli si scrive

$$A + B \xrightarrow{k} C + D \tag{1.2a}$$

in cui k è la costante di velocità, generalmente una funzione della temperatura e della pressione. A sinistra i *reagenti* A e B si combinano e scompaiono nel corso del tempo, mentre a destra i *prodotti* C e D vengono formati e appaiono con il procedere della reazione. Tuttavia in un sistema isolato si osserva che anche dopo molto tempo i reagenti A e B non scompaiono mai completamente. Piú precisamente, dopo un adeguato tempo di reazione, le quantità dei costituenti coesistenti A, B, C e D raggiungono un valore fisso del rapporto $c_C c_D / c_A c_B$, in cui c indica la concentrazione di ogni costituente. Quando si raggiunge questo valore fisso, si dice che il sistema si trova all'*equilibrio chimico*, e il valore del rapporto è la *costante di equilibrio*. Questo è l'analogo dell'omogeneo stato di quiete nel problema di Bénard. È chiaro che i valori dei parametri fisici quali pressione, temperatura, ecc. rimangono costanti.

Come si può conciliare questo risultato con l'equazione 1.2*a*, secondo la quale A e B sono obbligate a scomparire? Gli esperimenti mostrano l'esistenza di una trasformazione inversa della 1.2*a*

$$C + D \xrightarrow{k'} A + B \tag{1.2b}$$

All'equilibrio avvengono tutte due le reazioni con esattamente la stessa velocità. Questa fondamentale proprietà della natura, nota come *bilancio dettagliato,* è responsabile della maggior parte delle proprietà che caratterizzano questo tipo di fenomeni. Rappresentiamo cosí una reazione reversibile:

$$A + B \underset{k'}{\overset{k}{\rightleftharpoons}} C + D \tag{1.2c}$$

Nel problema di Bénard siamo stati in grado di allontanarci dall'equilibrio comunicando un flusso di energia al nostro sistema. L'analogo chimico piú ovvio è di sottoporre il sistema a un flusso di materia verso (o da) l'ambiente circostante, realizzando cosí ciò che verrà chiamato d'ora in poi un *sistema aperto*. Per esempio, si può eliminare C o D dal recipiente della reazione quando le loro concentrazioni diventano piú grandi di un valore prestabilito; oppure si può fornire al sistema un miscuglio ricco di A e B, che passa attraverso la reazione e che viene alla fine tolto o riciclato.

La figura 1.6 descrive il modo di operare di un sistema aperto in un laboratorio. Operando opportunamente sui valori di velocità di ingresso e di uscita del flusso di materia, si possono creare le condizioni affinché il sistema raggiunga uno stato in cui le concentrazioni di A, B, C, D rimangono costanti nel tempo, mentre nello stesso tempo il loro rapporto non è piú dato dalla costante di equilibrio. Matematicamente questo corrisponde all'azzerarsi delle variazioni delle concentrazioni nel tempo. Per esprimere questo in modo appropriato, introduciamo la derivata rispetto al tempo e scriviamo

$$\frac{dc_A}{dt} = \frac{dc_B}{dt} = \ldots = 0 \qquad (1.3)$$

Chiameremo lo stato sopra descritto uno *stato stazionario di non-equilibrio.*

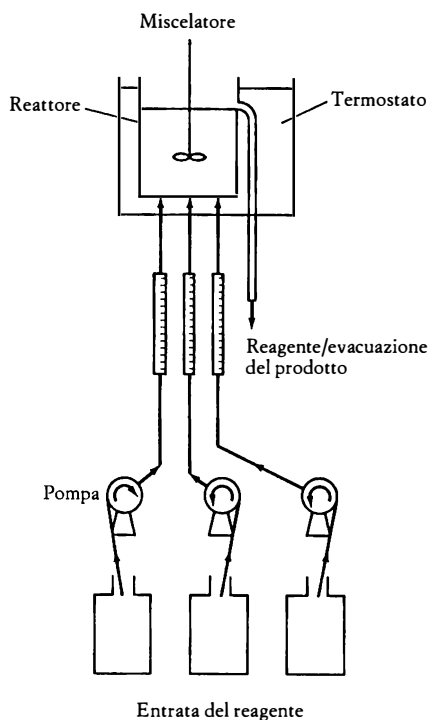
Possiamo aspettarci da un sistema aperto in un tale stato di non-equilibrio un comportamento simile al problema di Bénard? Per un certo verso, in un siffatto sistema non regge piú il bilancio dettagliato. Supponiamo ora che, momentaneamente, un particolare processo, per esempio il passo successivo di una reazione, aumenti molto in confronto al passo precedente. All'equilibrio, il bilancio dettagliato tenderebbe a ristabilire le condizioni iniziali. Lontano dall'equilibrio, tuttavia, non è piú cosí. Per di piú, se qualche parte del meccanismo chimico fosse in grado di mettere il sistema in grado di catturare e amplificare ulteriormente il sopra menzionato aumento, si potrebbe avere una situazione potenzialmente instabile simile a quella descritta per il problema di Bénard nella figura 1.4. Si sa che questi meccanismi esistono in chimica, e la loro manifestazione piú sorprendente è l'*autocatalisi.* Per esempio, la presenza di un prodotto può aumentare la velocità della propria produzione. È un dato di fatto che questo apparentemente esotico fenomeno accade normalmente in ogni processo di combustione, grazie alla presenza di radicali liberi, sostanze estremamente reattive che contengono un elettrone spaiato, e che, reagendo con altre molecole, dànno luogo a un ulteriore aumento di radicali liberi e a un processo di autoaccelerazione. E ancora, l'*autoriproduzione*, una delle maggiori proprietà caratteristiche della vita, è fondamentalmente un risultato di un ciclo autocatalitico, in cui il materiale genetico viene duplicato grazie all'intervento di specifiche proteine che vengono, a loro volta, sintetizzate dalle istruzioni contenute nello stesso materiale genetico.

Per molto tempo i chimici pensarono che da ogni trasformazione chimica dovesse alla fine scaturire uno stato omogeneo, indipendente

dal tempo, simile all'equilibrio. Ogni deviazione da questa regola poteva solo essere il risultato di un artificio, o una curiosità divertente che si realizzava a causa dell'interferenza con un fenomeno che non appartiene al regno della chimica.

Descriviamo ora una reazione chimica conosciuta come la reazione Belusov-Žabotinskij (d'ora in poi indicata con BZ), che, sotto certe condizioni di non-equilibrio, presenta una vasta gamma di affascinanti e inattesi comportamenti.

Non c'è nulla di particolare nei reagenti della reazione BZ. Una preparazione tipica consiste in solfato di cerio $Ce_2(SO_4)_3$, acido malonico, $CH_2(COOH)_2$ e bromato di potassio $KBrO_3$, disciolti in acido solforico. L'evoluzione del sistema può essere seguita a occhio nudo, poiché un eccesso di ioni Ce^{4+} determina un colore giallo chiaro, mentre un ec-

Figura 1.6
Realizzazione sperimentale di un sistema chimico aperto.

cesso di ioni Ce^{3+} conferisce un aspetto incolore. Mezzi di osservazione piú sofisticati vengono forniti da speciali elettrodi, o da misure spettroscopiche di assorbimenti ottici causati da una particolare sostanza.

Vediamo ora i tipi di comportamento mostrati da questo sistema sotto diverse condizioni sperimentali, tutte a temperatura normale.

Reazione BZ in sistemi ben mescolati: orologi chimici e caos.

Supponiamo che la reazione si svolga dapprima sotto le condizioni mostrate in figura 1.6. Grazie all'efficiente trasporto di materia assicurato dall'agitatore, il sistema rimane omogeneo nello spazio in ogni momento. Questo apparato sperimentale permette anche di controllare molto facilmente la distanza del sistema dall'equilibrio: è sufficiente pompare le sostanze chimiche dentro (o fuori) al sistema a diverse velocità, realizzando cosí diversi *tempi di permanenza* di una certa sostanza entro il recipiente della reazione. Un tempo di permanenza molto lungo equivale essenzialmente a realizzare un sistema chiuso, e sotto queste condizioni ci aspettiamo di raggiungere un comportamento simile all'equilibrio caratterizzato dal bilancio dettagliato. Diminuendo questo tempo, non permettiamo che avvenga un completo equilibrio fra i passi della reazione in avanti e indietro, e sotto queste condizioni ci aspettiamo che il sistema manifesti un comportamento di non-equilibrio. Questo è precisamente quello che fa vedere l'esperimento. Per lunghi tempi di permanenza si raggiunge uno stato omogeneo statico, nel quale le concentrazioni delle sostanze chimiche rimangono indipendenti dal tempo. Questo è il tipico stato a cui un chimico sperimentale è abituato, e condivide tutte le proprietà qualitative dell'equilibrio chimico. È l'analogo dello stato della conduzione termica realizzata nel sistema di Bénard quando viene applicata fra i piatti una debole differenza di temperatura.

Se ora diminuiamo il tempo di permanenza, ci troviamo di fronte a un tipo di comportamento completamente differente. All'improvviso il sistema viene colorato completamente di giallo chiaro, cosa che indica un eccesso di ioni Ce^{4+}. Dopo pochi minuti (o anche una frazione di minuto, a seconda delle circostanze) la soluzione diviene incolore, cosa che indica un eccesso di ioni Ce^{3+}. Il processo prosegue giallo, incolore, giallo, incolore, giallo ecc., in modo ritmico con periodo e ampiezza perfettamente regolari, che dipendono unicamente dai parametri del sistema e sono perciò intrinseci al sistema. Questa oscillazione, che misura il tempo attraverso una dinamica generata internamente, costituisce

un orologio chimico. La figura 1.7 descrive il tipico comportamento dell'orologio.

Perché queste oscillazioni dovrebbero essere degne di nota? Che cosa c'è di significativo nel comportamento ritmico di un orologio chimico? Non è forse l'equivalente del ben noto fenomeno fisico dell'oscillazione del pendolo?

La figura 1.8 mostra una spiegazione della profonda differenza fra i due tipi di oscillazione. La parte superiore mostra un pendolo senza attrito che vibra attorno alla direzione verticale con un angolo massimo di apertura θ_1 (l'ampiezza della funzione periodica $\theta(t)$ che descrive il valore istantaneo dell'angolo) e un periodo T_1. Essa mostra anche la varia-

Figura 1.7

Tracce potenziometriche di log Br^- e log Ce^{4+}/Ce^{3+} in funzione del tempo durante la reazione di Belusov-Žabotinskij. Concentrazioni iniziali $CH_2(COOH)_2$ = 0,032M, $KBrO_3$ = 0,063M, $Ce(NH_4)_2(NO_3)_5$ = 0,01 M, H_2SO_4 = 0,8 M e KBr = 1,5 × 10^{-5} M.

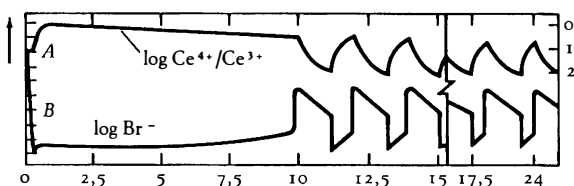
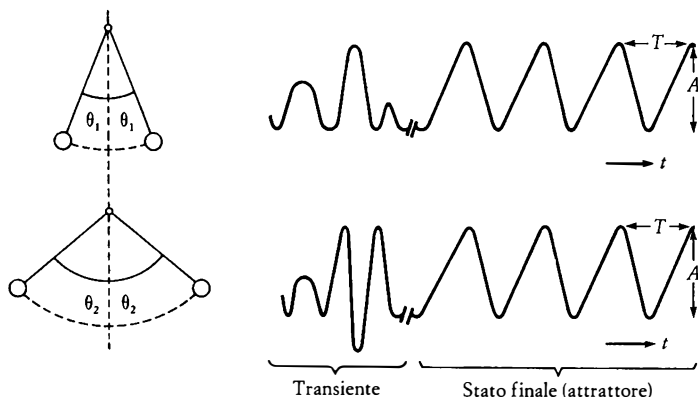



Figura 1.8

Confronto fra le oscillazioni forzate in un sistema conservativo come il pendolo (a sinistra) e nei sistemi dissipativi come i reagenti nella BZ (a destra).

zione nel tempo della concentrazione di una sostanza chimica nei reagenti della BZ, caratterizzata da una ampiezza *A* e da un periodo *T*.

Disturbiamo ora momentaneamente questo stato di cose, rispettivamente spostando il pendolo a un angolo θ_2 dalla verticale ($\theta_2 > \theta_1$), e applicando un leggero impulso di temperatura o di concentrazione nei reagenti della BZ (per esempio, iniettando poche millimoli di bromato di potassio o toccando il recipiente per qualche secondo).

La parte inferiore della figura 1.8 descrive la risposta dei sistemi a questa azione di disturbo. Il pendolo oscillerà ancora, ma la sua ampiezza sarà uguale a θ_2 invece che a θ_1, e il suo periodo sarà leggermente maggiore. In altre parole questo sistema ricorderà per sempre la perturbazione. Al contrario, dopo un effetto transiente, i reagenti della BZ si rimetteranno a oscillare esattamente con la stessa ampiezza e lo stesso periodo di prima. Questa è la proprietà della *stabilità asintotica* a cui abbiamo accennato nel paragrafo precedente. È direttamente collegata a una onnipresente proprietà della maggior parte dei fenomeni naturali, ossia *l'irreversibilità*, ed è essenzialmente responsabile della riproducibilità degli eventi. D'altra parte i sistemi come il pendolo, non hanno questa proprietà perché la loro dinamica è invariante sotto l'inversione del tempo. Sono dunque completamente alla mercé delle perturbazioni che possono investirli e, poiché queste perturbazioni sono essenzialmente impredicibili, essi sono costretti a mostrare prima o poi un comportamento irregolare.

Nel capitolo III presenteremo uno studio piú sistematico delle differenze fra i *sistemi dissipativi* che mostrano irreversibilità, come il flusso di Bénard o la reazione BZ, e i *sistemi conservativi*, come il pendolo, dove la dinamica è invariante sotto l'inversione temporale. Per ora un altro esempio mostrerà l'importanza dei sistemi dissipativi. Un sistema oscillatorio molto familiare che portiamo con noi fin dalla nostra vita embrionale, è il nostro cuore. Esso batte piú o meno regolarmente (circa 70 o 80 battiti per minuto per un individuo medio), ma può anche diventare irregolare. Una forma particolarmente drammatica di irregolarità è la fibrillazione, l'improvvisa incapacità di compiere la serie completa di azioni muscolari che costituiscono un intero battito cardiaco. (Qualunque sia la causa ultima del decesso di un individuo, la morte implicherà sempre uno stato di fibrillazione del cuore). Supponiamo che, a causa di una perturbazione, il modello di oscillazione – ossia il normale ritmo cardiaco – venga sconvolto. Dal momento che l'intero organismo umano è soggetto quotidianamente a un gran numero di perturbazioni, se il nostro cuore funzionasse come un pendolo la fibrilla-

zione potrebbe avvenire già nell'embrione, prima della nascita. Ma il cuore non è come un pendolo, esso non «ricorda» gli effetti di una perturbazione mutando permanentemente il modello di oscillazione: se non sopravviene alcun danno fisico e se la causa della perturbazione viene rimossa il cuore riprende il suo ritmo normale. Questo è vero per ogni altro fenomeno riproducibile osservato in natura, dai ritmi circadiani e i cicli di divisione delle cellule al cambiamento di luminosità delle stelle variabili, le cefeidi. Il fatto che questo e tutti gli altri fenomeni ritmici riproducibili appartengano allo stesso regno delle oscillazioni dei reagenti nella BZ, mostra l'enorme importanza dell'irreversibilità e dei sistemi dissipativi.

Torniamo ora al nostro orologio chimico. Nel vocabolario già introdotto nel § 3, si può dire che nel regime di stato stazionario (che è anche asintoticamente stabile) il sistema ignora il tempo. Ma una volta entrato nel regime periodico, esso «scopre» tutto a un tratto il tempo nella fase del moto periodico e nel fatto che i massimi delle diverse concentrazioni si susseguono l'una con l'altra in un ordine prescritto. Chiamiamo questo fatto *rottura della simmetria temporale*.

Da un punto di vista ancora piú fondamentale, il mantenimento del comportamento oscillatorio forzato che coinvolge l'intero sistema implica che le sue diverse parti agiscono di concerto, mantenendo una fase di relazione fra loro, altrimenti un'interferenza distruttiva potrebbe cancellare il comportamento oscillatorio. In altre parole, ci aspettiamo, come anche nel problema di Bénard, l'emergere di una *correlazione a lungo raggio* indotta dai vincoli del non-equilibrio. Nei capitoli II e IV verranno analizzate in dettaglio la produzione e le caratteristiche di queste correlazioni.

Ma non abbiamo ancora terminato la lista delle sorprese nascoste nella reazione BZ. Analisi sperimentali dettagliate, mostrano che per valori di tempo di permanenza che sono intermedi fra due tipi di regimi oscillatori, si osserva un comportamento complesso non periodico. La comparsa di questa *turbolenza chimica* illustra ancora una volta la tendenza di molti sistemi naturali a evolvere, sotto certe condizioni, in modo caotico. Essa fa anche vedere la seguente importante proprietà dei sistemi chimici: nella meccanica dei fluidi e in molti altri esempi fisici, il comportamento complesso è invariabilmente associato a disomogeneità spaziali. Ma in chimica anche un sistema che presenti omogeneità spaziale può mostrare nel tempo un comportamento complesso. La ragione è già stata citata all'inizio di questo paragrafo: i sistemi chimici sono dotati di meccanismi come l'autocatalisi, dovuti alla peculiare struttura molecolare e reattività di certi costituenti, che permettono lo-

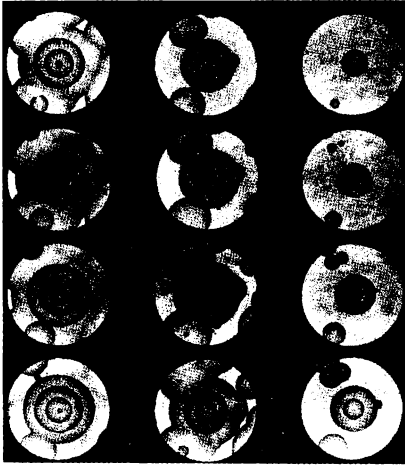
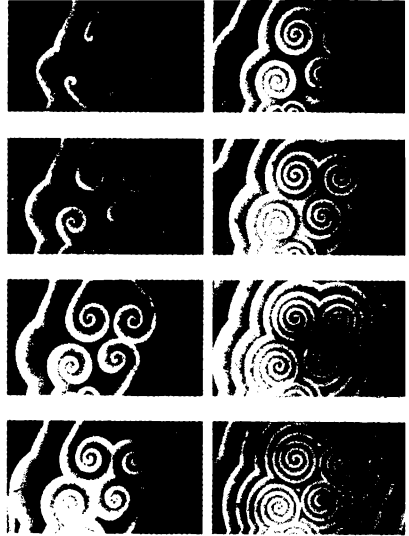
ro di evolvere verso nuovi stati amplificando (o reprimendo) gli effetti di deboli perturbazioni.

La reazione BZ in un sistema non uniforme: le strutture spaziali.

Supponiamo ora che la reazione venga portata avanti senza mescolare, permettendo cosí un possibile sviluppo di disomogeneità spaziali. Si osservano allora strutture regolari spazio-temporali nella forma di fronti d'onda che si propagano, la cui necessaria rappresentazione statica (figura 1.9) dà solo una debole idea della bellezza e dell'attività implicate. Le onde che si vedono nella figura sono create in un sottile strato di reagente. Esse appaiono di due diverse forme: fronti circolari (*a*), che mostrano un'approssimativa simmetria cilindrica attorno a un asse perpendicolare allo strato, generalmente chiamati *strutture a bersaglio*; e fronti a spirale (*b*) che ruotano nello spazio in senso orario o antiorario. È anche possibile ottenere, sebbene sotto eccezionali condizioni, le spirali a molte braccia, mostrate nella parte (*c*) della figura. In ogni caso i fronti d'onda trasportano, lungo distanze spaziali macroscopiche, senza distorsioni e a una velocità prescritta, il «messaggio» rilasciato dalla chimica nel centro da cui parte l'intero sistema. Ancora una volta, siamo testimoni della nascita della complessità.

Come nel problema di Bénard, possiamo associare la formazione dei fronti d'onda a una rottura spaziale di simmetria. La simmetria rotta dalle strutture a bersaglio della figura 1.9*a* è molto simile a quella rotta nel sistema di Bénard: essenzialmente, il sistema non è piú invariante alle traslazioni lungo una particolare direzione dello spazio. La rottura di simmetria legata alla formazione delle spirali (figura 1.9*b* e *c*) è, d'altra parte, molto diversa, poiché è associata alla nozione di *chiralità*, o rotazione. Questo tipo di asimmetria della materia ha sempre avuto un fascino particolare. Louis Pasteur, il fondatore della moderna biochimica, espresse ripetutamente il suo stupore di fronte alla asimmetria ottica delle biomolecole, che si manifesta con la rotazione del piano di polarizzazione della luce in una particolare direzione. Egli considerò questa proprietà come uno degli aspetti basilari della vita. Per di piú l'osservazione della asimmetria morfologica dell'organismo adulto ha introdotto nel pensiero umano la nozione di «destra» e «sinistra», che ha influenzato filosofi e scrittori fin dai tempi di Platone. È stupefacente vedere come queste idee profonde emergano naturalmente da una dinamica intrinseca di un modesto e apparentemente ordinario sistema.

Il comportamento sopra descritto non è un aspetto esclusivo del si-

Figura 1.9
Propagazione delle onde in uno strato bidimensionale di reagenti BZ.
a) Strutture a bersaglio.
b) Onde a spirale.
c) Spirali a molte braccia.

a)

b)

c)

stema BZ. Esso viene condiviso da una moltitudine di altri fenomeni che si realizzano nella fase omogenea e che coinvolgono ancora semplici sostanze chimiche, delle quali viene data una parziale lista nella tabella 1. Tipicamente, in tutte queste situazioni il dominio dei valori parametrici per cui le oscillazioni vengono osservate, è molto vicino al dominio nel quale avviene un altro interessante fenomeno: la *bistabilità*. Piú precisamente, due (o anche piú) stati simultaneamente stabili coesistono sotto le stesse condizioni sperimentali. Il particolare stato scelto dal sistema dipenderà dalle condizioni sperimentali. Questo è illustrato dalla figura 1.10, nella quale vi è il diagramma del valore di una variabile (ad es. la concentrazione di una sostanza chimica) rispetto a un parametro caratteristico di controllo (ad es. il tempo di permanenza). Supponiamo che un esperimento sia realizzato per un valore $\lambda = \lambda'$ di questo parametro, per il quale un solo stato stabile (il ramo *a*) sia utilizzabile. Aumentando gradualmente λ entriamo in una regione di molteplici sta-

Tabella 1.

Oscillatori chimici in un CSTR

Specie principali	Specie addizionali
I^-	IO_3^-, MnO_4^- o $Cr_2O_7^-$
I^-	Acido malonico
IO^{3-}	H_3AsO_3
IO^{3-}	$Fe(CN)_6^{4-}$, SO_3^{2-}, acido ascorbico o $CH_2O \cdot SO_2$
I_2	$Fe(CN)_6^{4-}$, SO_3^{2-}
IO^{3-}	I^-, acido malonico
IO^{3-}	I^-, H_3AsO_3
I^-	BrO_3^-
BrO_3^-	SO_3^{2-}, $Fe(CN)_6^{4-}$, H_3AsO_3 o Sn^{2+}
I^-	I_2, $S_2O_3^{2-}$

Oscillatori chimici nella catalisi superficiale

Specie principali	Specie addizionali
$CO + O_2$	Pt, Pd, CuO, Ir
$H_2 + O_2$	Pt, Pd, Ni
$NH_3 + O_2$	Pt
$C_2H_4 + O_2$	Pt
$C_3H_6 + O_2$	Pt
$C_6H_{12} + O_2$	NaY (zeolite)
N_2O (scomposizione)	CuO

ti ($\lambda_1 < \lambda < \lambda_2$). Il sistema rimane tuttavia nel ramo (*a*) finché λ supera il valore λ_2. A questo punto salta nel ramo (*b*) e lí rimane. Ora, se si impone una variazione di λ nella direzione opposta, partendo, per esempio, da un valore λ'' e andando verso il valore λ', il sistema rimarrà sul ramo (*b*) fino al valore λ_1 del parametro. In altre parole il sistema mostra strutture diverse a seconda della sua storia passata. Chiamiamo questo fenomeno *isteresi*.

In tutti i casi noti fino ad ora, risulta che lo stesso meccanismo chimico può spiegare sia la bistabilità che le oscillazioni e le onde. Come già detto, vi è generalmente coinvolto almeno un processo di autocatalisi. Nel sistema BZ questo processo assicura la produzione di due moli di una sostanza che appare come un intermediario nella reazione, l'acido bromoso ($HBrO_2$), per ogni molecola della stessa sostanza secondo la

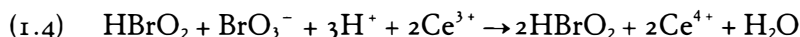
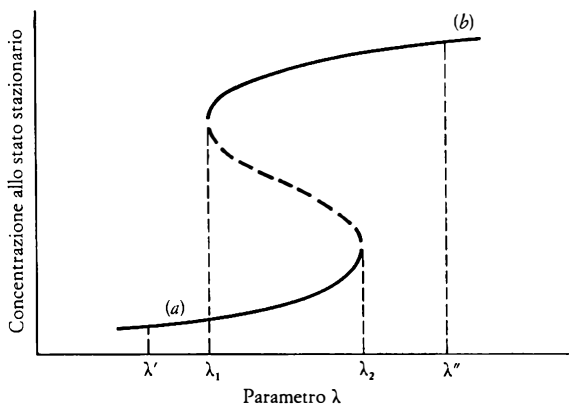
$$(1.4) \qquad HBrO_2 + BrO_3^- + 3H^+ + 2Ce^{3+} \rightarrow 2HBrO_2 + 2Ce^{4+} + H_2O$$

L'autocatalisi diretta non è tuttavia l'unico meccanismo capace di generare un comportamento complesso in chimica. Per esempio, nel fe-

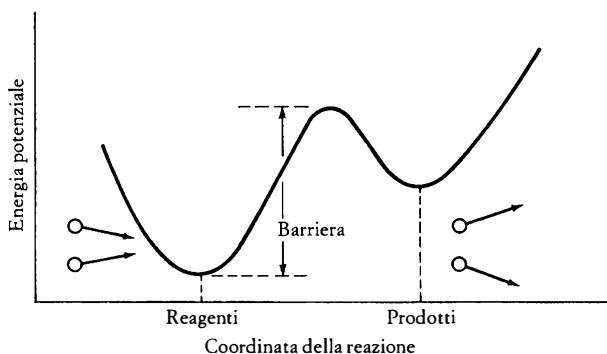
Figura 1.10

Illustrazione dei fenomeni della bistabilità e dell'isteresi. Il sistema è fatto partire nel ramo *a* con un valore iniziale del parametro λ uguale a λ'. Poi λ viene aumentato. Il sistema dapprima resta in *a*) ma oltre il valore del punto limite $\lambda = \lambda_2$, salta nel ramo *b*). Quando λ viene fatto decrescere al di sotto del valore λ'', il sistema rimane nel ramo *b*) finché λ diviene λ_1, punto in cui salta in *a*). In altre parole il percorso specifico degli stati del sistema dipende dalla sua storia passata. Ciò è noto come isteresi. Si noti che nella regione compresa tra $\lambda_1 < \lambda < \lambda_2$ il sistema presenta bistabilità, ossia la possibilità di evolvere, per dati valori del parametro, in piú di uno stato stabile, a seconda della preparazione iniziale.

nomeno della catalisi eterogenea, della quale abbiamo già parlato all'inizio di questo paragrafo, alcuni processi chimici liberano energia (sono perciò chiamati reazioni esotermiche). Questo porta a un riscaldamento del mezzo. Ora, come già detto, la costante di velocità di una reazione chimica dipende dalla temperatura. È in effetti una funzione crescente della temperatura, descritta con buona approssimazione dalla legge di Arrhenius:

$$k(T) = k_0 e^{-E_0/k_B T} \tag{1.5}$$

nella quale E_0 è l'energia di attivazione e la costante k_B è una costante universale chiamata costante di Boltzmann. Questa dipendenza può essere compresa qualitativamente da quanto segue, e una possibilità è mostrata nella figura 1.11. Una trasformazione reattiva implica la rottura di legami chimici, in altre parole deve superare una «barriera di energia» corrispondente all'energia di legame. Questa quantità di energia è fornita dall'energia cinetica del moto traslazionale delle molecole interagenti. Se il mezzo viene scaldato, l'energia cinetica media delle molecole aumenterà, e cosí un gran numero di coppie di molecole che si urtano possiederanno una quantità sufficiente di energia per superare la barriera. In altre parole la reazione verrà accelerata secondo l'equazione 1.5.

Figura 1.11

Una reazione chimica rappresentata in termini del moto di un potenziale a doppio pozzo. I due minimi corrispondono a configurazioni stabili delle specie coinvolte nel processo che sono identificate come *reagenti* e *prodotti*.

Supponiamo ora di avere a che fare con una reazione esotermica. Se questo processo viene accelerato, verrà liberato piú calore, la temperatura del mezzo aumenterà ulteriormente, e la reazione verrà ulteriormente accelerata. Questo fatto fornirà dunque un elemento potenzialmente destabilizzante, capace di indurre transizioni a nuovi tipi di comportamento. E infatti, in molte reazioni catalitiche di interesse industriale, vengono osservate strutture del tutto simili a quelle descritte per il sistema BZ. Queste incidono considerevolmente sullo stato locale del catalizzatore e giocano un ruolo molto importante nel corso della reazione, nel rendimento e nell'efficienza dell'impianto chimico.

5. *Complessità fisico-chimica e complessità algoritmica.*

Abbiamo incontrato ora tre modi fondamentali di autoorganizzazione della materia che dànno luogo a un comportamento complesso: la bistabilità e l'isteresi, le oscillazioni (sia periodiche che non periodiche), e le strutture spaziali. Abbiamo visto questi fenomeni emergere in due casi: il flusso del mezzo sotto un vincolo termico e un sistema aperto soggetto a reazioni chimiche catalitiche. Vediamo se, con l'aiuto di questi esempi, possiamo ottenere una chiara definizione di complessità. Ciò dovrebbe esserci di grande aiuto per la nostra successiva analisi dei meccanismi responsabili della comparsa di un comportamento complesso.

Un modo per arrivare a una definizione che colga l'essenza della complessità è fare un'astrazione dei dettagli degli specifici fattori sperimentali considerati nei §§ 3 e 4, e cercare di comunicare a un altro osservatore un insieme di istruzioni che lo mettano in grado di riprodurre il comportamento osservato senza eseguire di nuovo l'esperimento. Supponiamo che il costo per comunicare con un tale osservatore sia molto alto. Ciò introduce il vincolo addizionale di rendere le istruzioni piú corte possibile senza compromettere la realizzazione dello scopo iniziale.

Come dovremmo procedere in queste circostanze nel caso dell'esperimento di Bénard? Per descrivere la situazione di equilibrio illustrata nella figura 2 le seguenti istruzioni minimali sono sia necessarie che sufficienti[2]:

[2] Per semplicità concentreremo l'attenzione sui dati riguardanti le condizioni al contorno e la velocità v, lasciando da parte temperatura e pressione.

- $\Delta T = T_2 - T_1 = 0$
- $v = 0$ in ogni punto su un piano (orizzontale) parallelo ai piatti, dove v è la velocità di convezione.

Proviamo ora a comunicare gli aspetti alla struttura che sorge al di là della instabilità di Bénard. Con riferimento alle figure 1.3 e 1.5, conveniamo che il messaggio dovrebbe essere circa il seguente:

- $\Delta T = T_2 - T_1 > \Delta T_c$
- $v = v_0$ in un particolare punto P scelto come punto di riferimento
- Questo punto particolare appartiene a una cella che ruota in una ben definita direzione (S o D)
- v in ogni altro punto può essere dedotto da v_0 sulla base della sequenza periodica SDSD..., e dal valore della lunghezza d'onda della struttura.

Vediamo che come ci muoviamo attraverso la soglia di Bénard diventano necessarie piú istruzioni per specificare lo stato generato dal sistema. Ciò invoca automaticamente la definizione di *complessità algoritmica*, proposta indipendentemente verso il 1965 dal matematico sovietico A. Kolmogorov e dal matematico americano G. Chaitin. La complessità algoritmica di una sequenza di dati è definita dalla lunghezza minima di un algoritmo computazionale (misurata, per esempio, dal numero di bit se l'algoritmo deve essere comunicato a un computer digitale) che dovrebbe produrre questa sequenza in uscita.

La complessità algoritmica appena definita è equivalente alla complessità osservata nelle scienze fisiche e in biologia? Consideriamo una sequenza di dati di lunghezza N espressa in forma binaria e che mostra una regolarità complessiva (per esempio, 100 100 100...) Chiaramente, il messaggio in essa contenuto si può comprimere considerevolmente. Per esempio, si può trasmettere a un computer con un algoritmo molto semplice «Stampa 100 dieci volte (o 100 o un milione di volte)». Il numero di cifre binarie in un tale algoritmo è una piccola frazione del numero nella serie iniziale, e quando la lunghezza della serie diventa piú grande, la lunghezza dell'algoritmo cresce molto piú lentamente. Secondo la definizione, ciò implica una limitata complessità algoritmica. Consideriamo ora l'insieme di tutte le sequenze lunghe N cifre. Il numero di queste sequenze è 2^N. Supponiamo che quando N aumenta qualche sequenza può essere compressa come nell'esempio precedente,

e cioè, può essere espressa nella forma piú compatta di sequenze lunghe 1,2, ... K cifre, in modo che K/N tenda a zero quando N tende all'infinito. Il numero totale di tali sequenze «ordinate» è:

$$2 + 2^2 + \ldots + 2^K = 2^{K+1} - 2$$

Questo numero è molto piccolo in confronto a 2^N quando N diventa grande. In altre parole la maggior parte delle sequenze casuali diventerà incomprimibile. Secondo la definizione ciò implica una massima complessità algoritmica che è essenzialmente uguale alla lunghezza della sequenza N. Questo tipo di complessità si può generare in un tipico esperimento di lancio di monete, sempre che si prendano tutte le precauzioni per eliminare ogni correlazione possibile fra passi successivi, assicurando cosí una competizione «corretta».

Ora, questo è lontano dall'avere una corrispondenza con l'idea intuitiva suggerita dai risultati sperimentali dei §§ 3 e 4, in cui la complessità deve essere associata a una coerenza a lungo raggio e a un comportamento concordato. Vero è che l'ordine completo nella forma di una totale mancanza di variabilità è un caso estremo di coerenza, in cui l'oggetto è come un fossile, e difficilmente il suo comportamento può venire caratterizzato come complesso. D'altra parte, la forte variabilità rappresentata dal rumore casuale e dalla concomitante mancanza di correlazione è un'altra forma di organizzazione ugualmente non rappresentativa. Nella realtà la complessità fisico-chimica deve essere in qualche modo racchiusa fra questi due estremi, e non deve essere identificata completamente con la complessità algoritmica.

Come abbiamo visto, sotto certe condizioni i sistemi fisico-chimici possono generare casualità sotto forma di una dinamica caotica. Vedremo piú avanti, nei capitoli III, § 10, e IV, §§ 7 e 8, che ci sono alcune differenze fondamentali fra questa forma di casualità e la casualità generata dal rumore. Tuttavia, la dinamica caotica suggerirà alcune profonde connessioni fra la complessità fisica e algoritmica; queste non possono essere pienamente apprezzate allo stadio attuale della nostra discussione e verranno perciò posticipate al capitolo IV.

6. *Ulteriori esempi di comportamento complesso a livello macroscopico.*

Fino ad ora abbiamo incontrato tre modi fondamentali di autoorganizzazione della materia che dànno luogo a un comportamento complesso: la bistabilità e l'isteresi, le oscillazioni (sia periodiche che com-

plesse), e le strutture spaziali. Abbiamo visto questi fenomeni scaturire da due esempi che coinvolgono grossi flussi sotto un vincolo termico e sistemi aperti soggetti a reazioni chimiche catalitiche.

Come abbiamo visto, i tre modi di organizzazione discussi nei tre paragrafi precedenti si incontrano in un gran numero di altri problemi macroscopici, che scaturiscono in diversi contesti. Senza cercare di essere esaustivi, forniremo di seguito alcuni esempi, insistendo per ognuno sulle cause fondamentali che provocano la transizione.

Fenomeni indotti dalla tensione superficiale: la scienza dei materiali.

Lo scambio di energia e materia sotto condizioni di non-equilibrio tra identiche fasi della materia (solidi, liquidi, gas), tra fasi di diversi tipi, o tra membrane biologiche e liquidi o gas, ha luogo in molte circostanze. Questi processi sono caratterizzati dalla comparsa di *interfacce*, ovvero superfici di separazione fra le fasi presenti. Cosí come la pressione è il coefficiente di proporzionalità fra il lavoro fatto per cambiare il volume di un corpo e la variazione stessa del volume, la *tensione superficiale* è il coefficiente di proporzionalità fra il lavoro necessario per cambiare l'area di un elemento di superficie e l'area di questo elemento. Di regola, la tensione superficiale dipende dalla curvatura, dalla temperatura e dalla composizione. Costituisce dunque un meccanismo di accoppiamento fra queste quantità e le proprietà meccaniche del materiale.

Una manifestazione di questo accoppiamento è la spontanea deformazione della superficie che dà luogo a strutture spaziali regolari sulla superficie stessa, a una massiccia fluttuazione nei fluidi adiacenti prima a riposo, e anche alla formazione di gocce. Questi fenomeni sono importanti in farmacologia, nell'industria alimentare o nel recupero dell'olio terziario.

Si osserva un'altra sorprendente manifestazione nella crescita di un solido puro o in una lega dalla fusione, quando l'avanzamento dell'interfaccia viene indotto da un fronte congelante, guidato dall'esterno, che si allontana dalla fusione. Oltre all'importanza industriale, questo processo si attua anche quando dalle gocce d'acqua si forma la neve. In tutti questi casi si osserva una varietà di strutture spaziali, come la formazione di lamelle o di dendriti, che hanno grandezze caratteristiche di molti ordini maggiori di una qualsiasi lunghezza cristallografica.

Collegata a questi due argomenti, vi è anche una recente e promettente applicazione dei fenomeni del non-equilibrio alla geologia. In

molti depositi geologici si osservano spettacolari strutture di mineralizzazione regolare in una vasta gamma di dimensioni spaziali (vedi figura 1.12): strati metamorfici (con spessori varianti da un millimetro a un metro), graniti (dell'ordine del centimetro), agate (da un millimetro a un centimetro), ecc. L'interpretazione tradizionale attribuisce queste strutture a fenomeni «sequenziali», conseguenti all'effetto di successivi cambiamenti di stagione o di clima. Sembra tuttavia che una interpretazione piú soddisfacente sia quella di attribuire queste strutture alle transizioni di rottura di simmetria, indotte da vincoli di nonequilibrio. Se questo punto di vista venisse confermato, verrà profondamente alterata l'interpretazione di numerosi depositi geologici.

In aggiunta a questi esempi, citiamo anche problemi di formazione e propagazione di difetti e fratture che coinvolgono, in un modo o in un altro, la transizione mediata dalla presenza di un'interfaccia.

Fenomeni cooperativi indotti da effetti elettromagnetici: circuiti elettrici, laser, bistabilità ottica.

L'ingegneria elettrica e la scienza informatica fanno largo uso dei fenomeni cooperativi che sorgono in sistemi coinvolgenti elementi «attivi» come i diodi Gunn, i circuiti risonanti applicati a un triodo, o i semiconduttori. Dopo tutto, il funzionamento di ogni computer di oggi-

Figura 1.12

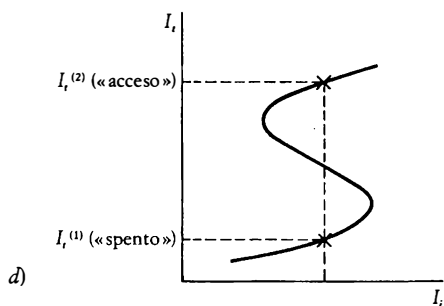
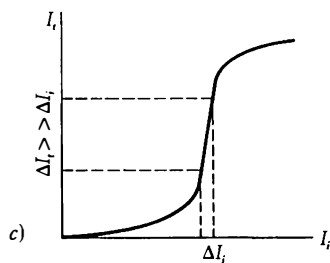
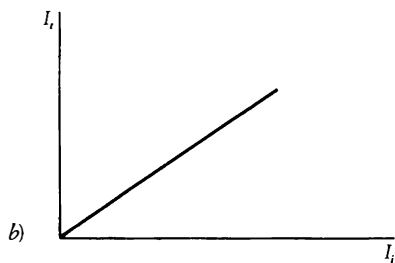
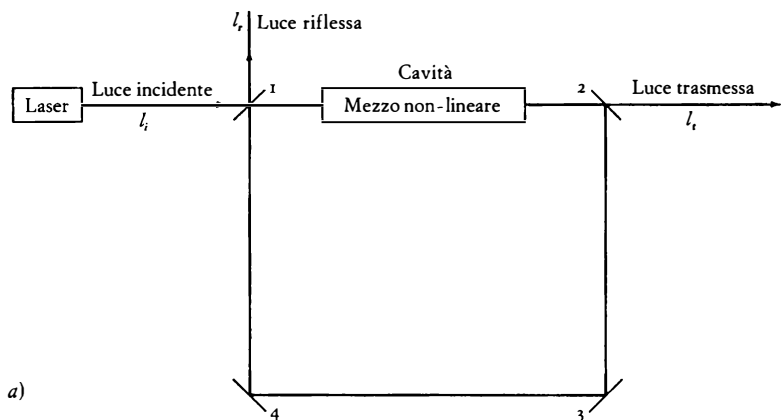
A sinistra: *skarn* da San Leone, Sardegna. Le linee chiare, spesse da 1 a 2 mm, consistono di granatoandranditico (calcio e ione ferrico). Le linee scure, spesse da 5 a 8 mm sono di magnetite e quarzo. Il rettangolo bianco è lungo 1 cm (per gentile concessione di B. Guy).
A destra: diorite orbicolare da Epoo, Finlandia. Gli strati concentrici sono alternativamente piú ricchi in biotite (scuri) e in plagioclase (chiari). Il raggio dell'orbicolare è 10 cm (per gentile concessione di E. Merino).

giorno implica circuiti con elementi bistabili, capaci di commutare fra due diversi stati secondo i valori di certi parametri. Dagli anni '60 si è avuta nell'ottica una vera rivoluzione, in una materia fino a quel tempo reputata una delle piú classiche e ben assestate branche della fisica. La produzione di fasci di luce coerente dai laser, originò impressionanti sviluppi sperimentali e teorici che stabilirono quanto questo fenomeno possa essere visto come un'autoorganizzazione che nasce da condizioni di non-equilibrio. Per di piú, la bistabilità ottica, la possibilità di avere due modi stabili di operazione in una cavità risonante, pone la tecnologia moderna di fronte all'affascinante sfida di costruire computer basati su circuiti bistabili elettro-ottici, che potrebbero essere piú vantaggiosi di quelli elettronici per i piú brevi tempi di commutazione e per il minor consumo di energia.

Riassumiamo brevemente il meccanismo responsabile della cooperatività e delle transizioni nel caso particolare della bistabilità ottica.

Supponiamo che un campo coerente di energia elettromagnetica, avente una frequenza nettamente definita nello spettro visibile – cioè un raggio di luce coerente –, venga immesso in una cavità (figura 1.13*a*). Questi campi sono generati molto facilmente dai moderni laser. La cavità viene sistemata in modo da essere nelle vicinanze del campo incidente. Ovviamente, quando essa è vuota, l'intensità del campo trasmesso I_t è proporzionale all'intensità incidente I_i (figura 1.13*b*). Tuttavia sorgono effetti del tutto inaspettati quando la cavità viene riempita con un mezzo capace di assorbire la luce, che può essere anche quasi risonante con la luce incidente. Per cominciare l'interazione luce-materia è insitamente non-lineare. Per esempio, la velocità di cambiamento della polarizzazione elettrica del campione nel tempo, dipende sia dalla luce incidente, sia dal numero di atomi che sono in uno stato eccitato. Inoltre, ed è ancora piú importante, in una cavità ottica la disposizione degli specchi – indicati con 1, 2, 3 e 4 nella figura 1.13*a* – rimanda indietro al campione parte della luce che potrebbe altrimenti sfuggire nello spazio esterno. Ciò è equivalente a un *feedback* positivo molto simile all'autocatalisi chimica. Sotto certe condizioni questo *feedback* viene riflesso secondo il comportamento mostrato nella figura 1.13*c*: una regione della curva di trasmissione sviluppa «un guadagno differenziale» nel senso che la pendenza I_t contro I_i è maggiore dell'unità. Questo sistema lavora come un *transistore ottico*, poiché una lenta modulazione di intensità imposta al raggio incidente viene amplificata all'uscita. Sotto altre condizioni, si osserva che la curva di trasmissione dello stato stazionario assume una forma a «S» (figura 1.13*d*). Si può allora trovare una

Figura 1.13

a) Rappresentazione schematica di un apparato sperimentale che dà luogo a bistabilità ottica.
b) Intensità della luce trasmessa in funzione dell'intensità della luce incidente in una cavità vuota.
c) e *d*) Una cavità riempita con un mezzo otticamente attivo può funzionare come un transistore *c*) o come interruttore ottico *d*).

gamma di valori dell'intensità della luce incidente per la quale il sistema mostra un comportamento bistabile. I due livelli di trasmissione possono essere messi in corrispondenza agli stati logici «0» e «1», suggerendo cosí la possibilità che un simile dispositivo possa essere la base di una *memoria ottica*.

7. *Ancora sui sistemi biologici.*

Non possiamo andare avanti nella nostra indagine sui comportamenti complessi senza spendere un po' di tempo per discutere i sistemi biologici in modo piú sistematico, andando oltre le scarse osservazioni intuitive fatte piú sopra in questo capitolo.

I sistemi viventi sono indubbiamente gli oggetti piú complessi e organizzati che si trovano in natura, dal punto di vista sia della loro morfologia che del loro funzionamento. Come già abbiamo detto, essi servono come prototipi dai quali le scienze fisiche prendono motivazioni e ispirazioni per comprendere la complessità. Sono letteralmente delle strutture storiche, poiché hanno l'abilità di preservare la memoria delle forme e delle funzioni acquisite nel passato, durante il lungo periodo di evoluzione biologica. Per di piú, i sistemi viventi funzionano certamente in condizioni lontane dall'equilibrio. Un organismo considerato nella sua interezza, riceve continuamente flussi di energia (per esempio il flusso solare usato dalle piante per la fotosintesi) e di materia (nella forma delle sostanze nutritive), che trasforma in diversi prodotti di rifiuto scaricati nell'ambiente. A livello cellulare, vengono osservate forti eterogeneità. Per esempio la concentrazione degli ioni potassio, K^+, entro le cellule del sistema nervoso, i neuroni, è piú alta di quella dell'ambiente esterno, mentre è vero il contrario per gli ioni sodio Na^+. Queste disuguaglianze, che implicano stati fortemente di non-equilibrio, sono all'origine di processi che giocano un importante ruolo nella vita, come la conduzione dell'impulso nervoso. Essi vengono mantenuti grazie al trasporto attivo e a reazioni bioenergetiche come la glicolisi e la respirazione.

Poiché siamo fin d'ora convinti che i sistemi fisico-chimici ordinari possono mostrare comportamenti complessi aventi molte caratteristiche usualmente attribuite alla vita, è legittimo chiedersi se qualcuno degli aspetti sopra citati dei sistemi biologici possa essere attribuito a transizioni indotte da vincoli di non-equilibrio e da appropriati meccanismi destabilizzanti simili alla autocatalisi chimica. Questa è probabil-

mente una delle questioni fondamentali della scienza. Non si può esigere alcuna risposta esaustiva, ma si possono menzionare alcuni esempi nei quali la connessione fra l'autoorganizzazione fisico-chimica e l'ordine biologico è piuttosto rilevante. Il problema su cui porteremo ora l'attenzione è il controllo dello sviluppo embrionale.

Lo sviluppo embrionale è la sequenza di eventi che porta da un'unica cellula, l'uovo fecondato, a un organismo completo. La natura ci offre un numero illimitato di esempi di tali processi. Il caso piú semplice è certamente quello dei batteri, il cui sviluppo si riduce a una sequenza di divisioni cellulari. All'altro capo dello spettro vi sono gli organismi superiori, come i mammiferi, dove lo sviluppo porta a un corpo pluricellulare in cui le cellule formano speciali tessuti e organi che possono comprendere anche 10^{12} o piú cellule.

Al momento attuale, è inimmaginabile arrivare a una comprensione dettagliata di come questi processi avvengano e, in particolar modo, di come sono coordinati con una precisione fantastica, che permette a ogni cellula di adempiere al suo ruolo al momento giusto e al posto giusto. Discuteremo piuttosto i sistemi viventi il cui sviluppo è caratterizzato da un livello «intermedio» di complessità, come l'ameba della specie *Dictyostelium discoideum*. Qui lo sviluppo si riduce essenzialmente a un fenomeno di transizione, molto simile a quello osservato nella reazione BZ, che segna il passaggio dallo stadio di vita unicellulare a quello pluricellulare.

La figura 1.14 descrive il ciclo di vita di questa specie. Nella (*a*) le amebe sono nello stadio unicellulare. Si muovono nel mezzo circostante, si nutrono di batteri e proliferano per divisione cellulare. Considerate globalmente esse costituiscono un sistema uniforme, poiché la loro densità (numero di cellule per centimetro quadrato) è essenzialmente costante. Supponiamo ora che le amebe siano soggette a inedia (in laboratorio questo viene ottenuto deliberatamente, in natura ciò può avvenire per condizioni ambientali meno favorevoli). Questo è analogo a un «vincolo» nell'esperimento chimico o fisico. Si osserva che le singole cellule non muoiono. Piuttosto esse rispondono al vincolo aggregandosi verso un centro di attrazione (figura 1.14*b*). L'iniziale omogeneità viene rotta, lo spazio diventa strutturato. Il risultante corpo pluricellulare è in grado di muoversi, presumibilmente per cercare condizioni di temperatura e di umidità piú favorevoli (figura 1.14*c*). Durante questa migrazione esso si differenzia e dà luogo a due tipi di cellule, una delle quali costituisce il peduncolo e l'altra il corpo fruttifero entro il quale si formano le spore (figura 1.14*d*). Alla fine, le spore vengono disseminate

nell'ambiente (figura 1.14*e*), e se le condizioni sono favorevoli, germinano, diventano amebe e il ciclo vitale inizia di nuovo.

Esaminiamo piú in dettaglio lo stadio di aggregazione mostrato dalla figura 1.15. Durante il processo si osservano i seguenti fenomeni. Prima, dopo la denutrizione, qualche cellula comincia a sintetizzare e a liberare nel mezzo extracellulare segnali di una sostanza chimica conosciuta come adenosinmonofosfato ciclico (cAMP). La sintesi e l'emissione sono periodiche, proprio come nell'orologio chimico del sistema BZ, con un periodo ben definito per fissate condizioni sperimentali. Diffondendo nel mezzo extracellulare, il cAMP emesso dalla cellula «pioniera» raggiunge la superficie delle cellule vicine. Si avviano allora due tipi di eventi. Primo, queste cellule realizzano un movimento orientato, chiamato *chemiotassi*, verso le regioni a piú alta concentrazione di cAMP, cioè verso la cellula pioniera. Questo moto dà luogo a strutture di densità delle cellule che assomigliano molto alle strutture ondulatorie dei reagenti BZ (figura 1.9). Secondo, il processo di aggregazione viene accelerato dall'abilità delle cellule sensibilizzate ad amplificare il segnale e di *ritrasmetterlo* al mezzo. Questo permette all'organismo di controllare un vasto territorio e di formare corpi multicellulari che comprendono qualcosa come 10^5 cellule.

Riassumendo, la risposta al vincolo dell'inedia dà luogo a un nuovo

Figura 1.14
Il ciclo vitale dell'ameba *Dictyostelium discoideum*.

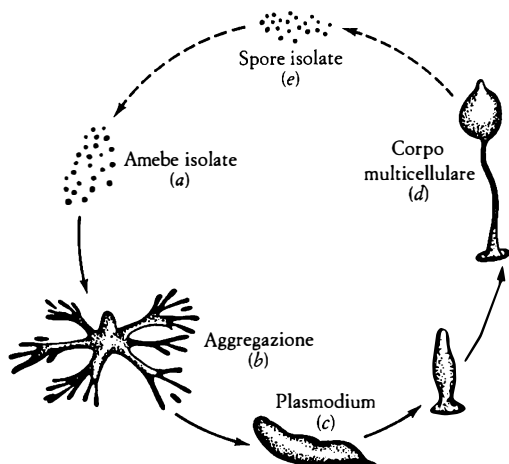
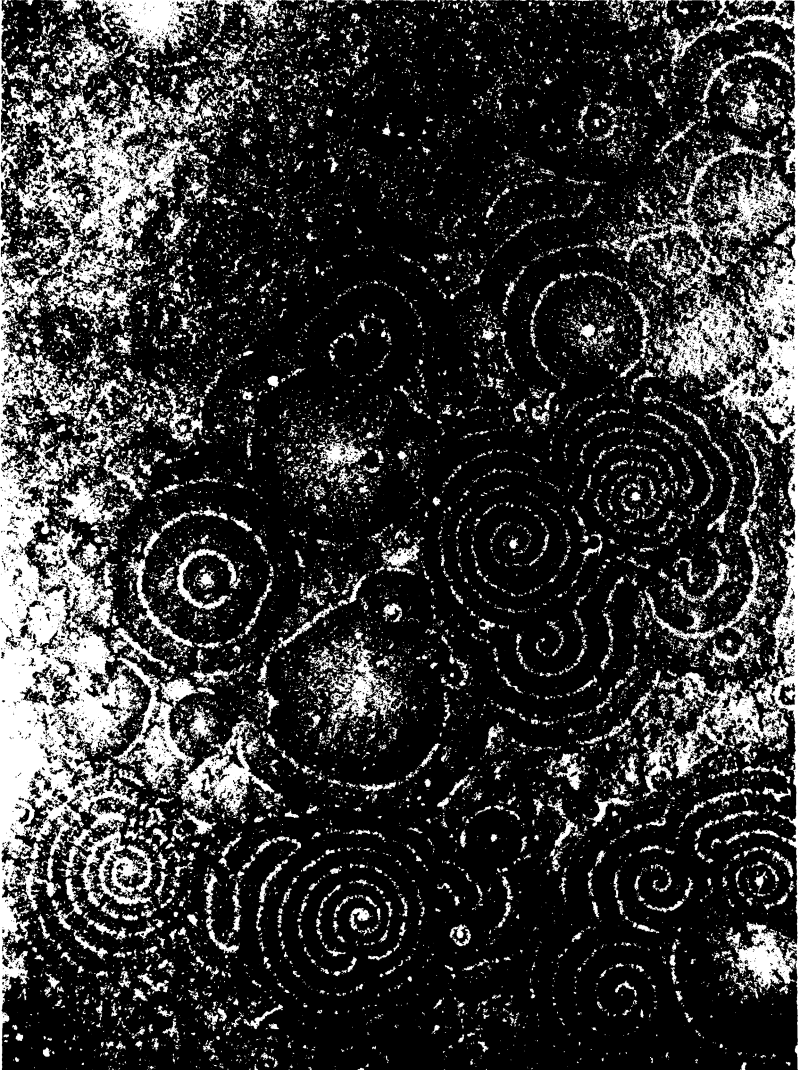



Figura 1.15

Onde concentriche e a spirale formate da popolazioni cellulari in aggregazione di *Dictyostelium discoideum* su una superficie di agar. I gruppi di amebe che si muovono verso il centro appaiono chiari e quelli stazionari scuri.

livello di organizzazione, che implica il movimento concordato di un gran numero di cellule e che permette all'organismo di rispondere flessibilmente a un ambiente ostile. Quali sono i meccanismi che mediano questa transizione? Osserviamo dapprima che il processo di chemiotassi porta a una amplificazione delle eterogeneità formate all'inizio, quando le cellule «pioniere» cominciano a emettere impulsi di cAMP. Poiché aumenta la densità delle cellule vicino al centro di emissione la chemiotassi contribuisce, grazie al fenomeno di ritrasmissione, a una maggiore emissione di cAMP da questa regione, aumentando cosí il movimento chemiotattico delle altre cellule verso di essa. Questo costituisce ciò che si chiama usualmente un «ciclo di *feedback*», molto simile all'autocatalisi chimica o alla autoaccelerazione di una reazione esotermica incontrate nel paragrafo 1.4.

Nel *Dictyostelium discoideum* si osserva un secondo meccanismo di *feedback* che opera al livello sottocellulare ed è responsabile sia dell'emissione periodica del cAMP sia della ritrasmissione del segnale chemiotattico. Questo meccanismo è collegato alla sintesi del cAMP nella cellula. Questa sostanza nasce dalla trasformazione di un altro importante costituente cellulare, l'adenosintrifosfato (ATP), che è uno dei principali trasportatori di energia entro le cellule viventi grazie al suo legame fosfato ad alta energia. La trasformazione ATP→cAMP non è tuttavia spontanea: è necessario un catalizzatore per accelerarla a un livello compatibile con i requisiti vitali. Nei sistemi biologici il ruolo di catalizzatori è assunto da speciali molecole, gli *enzimi*. Queste molecole contengono usualmente parecchi atomi e hanno un alto peso molecolare; da qui il termine *macromolecole* con cui vengono normalmente chiamati. Alcuni enzimi hanno un'unica zona attiva che i reagenti devono colpire per trasformarsi in prodotti. Ma in molti casi si ha a che fare con *enzimi cooperativi*, che dispongono di molti siti, alcuni dei quali sono catalitici e altri sono regolatori. Quando speciali molecole effettrici (o «modulatrici») si legano a questi ultimi siti, la funzione catalitica viene notevolmente influenzata. In alcuni casi le molecole agenti con la (o prodotte dalla) zona catalitica possono anche comportarsi come molecole effettrici. Questo darà il via a un ciclo di *feedback*, che sarà positivo (attivazione) se il risultato è un aumento della velocità della catalisi, o negativo (inibizione) in caso contrario.

L'enzima che catalizza la conversione ATP→cAMP è chiamato adenilico ciclasi ed è fissato alla membrana della cellula. Esso interagisce con un recettore presente sulla parte esterna della membrana in modo cooperativo, ma i dettagli della sua azione non sono ancora perfetta-

mente compresi. Una ovvia congettura è che un sito attivo dentro la cellula catalizzi la reazione. Il cAMP diffonde nel mezzo extracellulare attraverso la membrana della cellula e, con una certa probabilità, può legarsi alla parte esterna (regolatrice) dell'adenilico ciclasi e attivarla (figura 1.16).

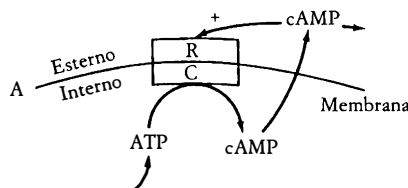
In questo modo aumenta la propria produzione, dando luogo a un ciclo di *feedback* capace di amplificare segnali e di indurre un comportamento oscillatorio.

I processi che segnano lo sviluppo del *Dictyostelium dicoideum*, per quanto semplici possano sembrare, sono il prototipo di una grande classe di fenomeni evolutivi piú sofisticati. Sembra infatti che la rottura della simmetria spaziale e la concomitante divisione in scomparti delle cellule, spieghi molti dati sperimentali della morfogenesi degli insetti. Piú generalmente, nello sviluppo dei tessuti si osservano frequentemente gradienti di varietà di sostanze come ioni o metaboliti. È naturale ipotizzare che questi gradienti procurino al tessuto un «sistema coordinato» che porta una *informazione posizionale* alle singole cellule, grazie alla quale esse possono riconoscere la loro posizione rispetto alle altre. È dunque piú probabile che le transizioni mediate da sostanze chimiche e che portano alla rottura di simmetria siano uno degli aspetti chiave della vita riconducibile ai fenomeni fisico-chimici autoorganizzati. Questa sorprendente idea fu annunciata per la prima volta nel 1952 dal matematico inglese Alan Turing ed è stata fin da allora una costante fonte di ispirazione sia per fisici che per biologi.

8. *La complessità a livello planetario e cosmico.*

Siamo ancora lontani dall'aver definito la complessità ed è improbabile che raggiungeremo lo scopo alla fine di questo saggio poiché la

Figura 1.16

Sintesi del meccanismo oscillatorio del cAMP nel *Dictyostelium discoideum*.

complessità è una di quelle idee la cui definizione è parte integrante dei problemi che essa solleva. Ciò che sappiamo fino ad ora, grazie ai numerosi esempi illustrati nei paragrafi precedenti, è che uno degli aspetti essenziali del comportamento complesso è la capacità di compiere *transizioni* fra diversi regimi. Detto diversamente, la complessità riguarda i sistemi nei quali l'evoluzione, e dunque la storia, gioca (o ha giocato) un ruolo importante nel comportamento osservato.

Quanto generali sono questi sistemi? Per molto tempo la scienza ha conservato la visione statica secondo la quale il mondo è una macchina ben oliata che lavora con la stessa implacabile precisione che lo caratterizzò nel lontano passato e che lo caratterizzerà ancora in un qualsiasi prevedibile futuro. Oggi questa visione può difficilmente essere sostenuta. La rivoluzione biologica iniziata da Darwin nel XIX secolo, e la rivoluzione che sta compiendo in geofisica e in cosmologia fin dagli anni '50 ci insegnano che dovunque guardiamo nel nostro ambiente, troviamo resti del passato i quali, lontani dall'essere puri fossili, stanno attivamente preparando la storia a venire. L'atmosfera della Terra come la conosciamo oggi è un risultato dello sviluppo della vita sul pianeta. Gli esseri viventi sono i risultati di una lunga evoluzione il cui stadio chimico e prebiotico venne direttamente influenzato dalle condizioni prevalenti sulla Terra all'incirca quattro miliardi di anni fa. I ghiacciai della Groenlandia o dell'Antartico, la posizione dei continenti, il fondo dell'oceano, il nostro clima, sono tutti stati soggetti a una lunga evoluzione caratterizzata da una serie di transizioni su larga scala. Ultimo esempio, ma non meno importante, l'universo come un tutto è in continua evoluzione, e la radiazione di fondo (equivalente a un corpo nero che irradia a 2,7°K) scoperta dagli astronomi, è un segno convincente della esplosione primordiale, il Big Bang, che creò l'universo circa 20 miliardi di anni fa.

Cominciamo ora a immaginare quanto lontano possano portare i fenomeni di transizione, simili a quelli che hanno luogo in un modesto recipiente di liquido o in una soluzione chimica. Non possiamo naturalmente esaurire qui l'argomento, sicché sceglieremo come caso esemplare il problema del cambiamento di clima, sul quale torneremo nel paragrafo successivo.

Dal 1960 circa un numero sempre maggiore di osservazioni porta alla conclusione che il clima della Terra mostra una variabilità intrinseca molto pronunciata. Questa conclusione, venendo dopo l'anormale periodo climatico favorevole e stabile che l'umanità ha conosciuto durante la prima parte del nostro secolo, produce sia stupore sia preoccupa-

zione negli specialisti, nei politici e nel pubblico in generale. Per la prima volta l'uomo comprende il carattere globale, planetario, del sistema climatico, come anche il fatto che le sue attività possono influire sul funzionamento della poderosa macchina del clima.

Vediamo ora un certo numero di fatti che hanno contribuito a questo nuovo atteggiamento, iniziando con il problema dei mutamenti *a lungo termine*. La maggior parte delle osservazioni a disposizione è il risultato della potente tecnica dell'analisi della composizione isotopica dei resti organici, che dà una diretta informazione sulle antiche temperature.

Uno dei primi punti che appare ovvio da questi dati è che le condizioni climatiche dominanti negli ultimi due o trecento milioni di anni erano estremamente diverse da quelle di oggi. Durante questo periodo, a eccezione dell'èra Quaternaria (la nostra èra, che ha inizio circa 2 milioni di anni fa), non vi era praticamente ghiaccio sui continenti e il livello dei mari era piú alto di quello odierno di circa 80 metri. Il clima era particolarmente tiepido, e le differenze di temperatura fra le regioni equatoriali (da 25 a 30°C) e polari (da 8 a 10°C) relativamente piccole.

È durante l'èra Terziaria, circa 40 milioni di anni fa, che iniziò a svilupparsi un contrasto fra le temperature equatoriali e quelle polari. Cosí nel tempo relativamente breve di centomila anni, la temperatura del mare a sud della Nuova Zelanda scese di parecchi gradi. Questo fu probabilmente l'inizio della corrente antartica, che indebolisce lo scambio di calore fra alte e basse latitudini e contribuisce a un ulteriore raffreddamento delle masse di acqua «intrappolate» in questo modo vicino alle regioni polari. Ancora una volta, vediamo in azione un meccanismo di *feedback*.

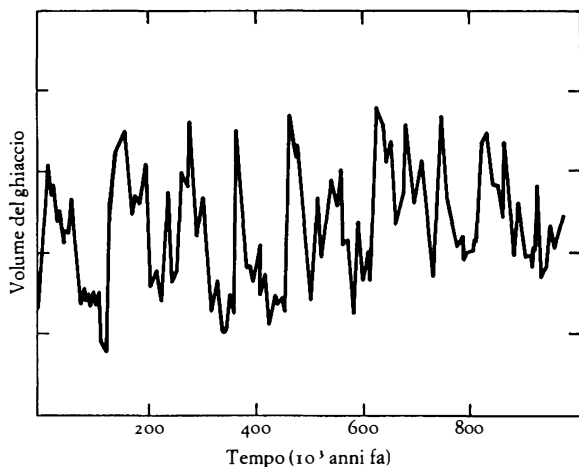
All'inizio dell'èra Quaternaria questa differenza era sufficientemente importante da permettere la formazione e il mantenimento del ghiaccio continentale. Nell'emisfero settentrionale ebbe luogo una serie di *glaciazioni*, che spinsero talvolta i ghiacci fino alle medie latitudini. Questi episodi climatici presentano una periodicità di circa centomila anni, alla quale viene sovrapposta un'apprezzabile quantità di casualità di disturbo (cfr. figura 1.17).

L'ultimo avanzamento del ghiaccio continentale nell'emisfero settentrionale raggiunse il suo massimo circa 18 000 anni fa, e le sue tracce sono ancora visibili. Oggi la quantità di ghiaccio continentale è circa 30 milioni di chilometri cubi (confinati essenzialmente nell'Antartico e nella Groenlandia); allora erano circa 70-80 milioni, che coprivano, in aggiunta, buona parte delle regioni settentrionali dell'America e del-

l'Europa. A causa della grande quantità di acqua imprigionata nei ghiacciai il livello del mare era piú basso di 120 metri. Da allora una gran parte del ghiaccio si è sciolta, disegnando le linee costiere e la maggior parte delle altre caratteristiche del paesaggio cosí come le conosciamo oggi.

Discutiamo ora i cambiamenti climatici in una scala di tempo intermedia. Le ultime tracce di calotte glaciali continentali (a eccezione della Groenlandia e dell'Antartico), scomparvero circa 7000 anni fa, e il periodo susseguente a questo evento è chiamato *optimum climatico*. L'aspetto piú sorprendente di questa èra è la comparsa di periodi umidi nella zona fra la Mauritania (17°O) e Rajasthan (78°E), oggi una delle zone piú aride del pianeta. Perfino nel centro del Sahara, che oggi riceve a malapena 5 mm di piogge all'anno, vi erano fiumi permanenti, una agricoltura avanzata, e dunque precipitazioni di circa 300-400 mm. Questa situazione favorevole sembra essere durata fino all'inizio dell'Età del Ferro, circa 3500 anni fa. Lo stile di vita delle popolazioni nel bacino orientale del Mediterraneo che ci rivelano i poemi di Omero o lo scalpello di Schliemann ci fa pensare a una perenne primavera, una natura generosa e una terra fertile.

Figura 1.17

Variazione del volume del ghiaccio continentale durante l'ultimo milione di anni, dedotto dalle datazioni isotopiche del centro dei mari profondi.

Da circa 3000 anni si osserva un raffreddamento globale. Tuttavia, questa tendenza generale viene interrotta da numerose fluttuazioni dell'ordine del secolo, che corrispondono a variazioni di circa 1 °C. Un ben noto esempio è il periodo dall'800 al 1200, durante il quale i Vichinghi scoprirono la Groenlandia (la «terra verde») e probabilmente l'America del Nord. Una tendenza opposta venne osservata fra il 1550 e il 1700. Questa «piccola glaciazione» costò all'Europa e all'America settentrionale carestie e molte altre catastrofi.

Comprendiamo ora come il clima della prima parte del nostro secolo va considerato piú un'anomalia che un esempio tipico nella lunga e tumultuosa storia climatica. Per alcuni decenni abbiamo avuto l'esperienza di un clima relativamente prevedibile, e abbiamo assistito al ritiro della calotta polare dell'emisfero settentrionale. L'agricoltura e la produzione alimentare ne hanno beneficato considerevolmente, e ciò ha contribuito all'aumento della popolazione mondiale. D'altra parte, questa apparente permanenza ha portato a idee errate su ciò che è o non è «normale» in climatologia.

Dalla metà degli anni '70 si osserva un ritorno alla variabilità del clima. Un esempio è dato dall'inverno eccessivamente rigido che colpí le regioni orientali dell'America del Nord nel 1976-77 e la simultanea prolungata siccità in quelle occidentali. Sembra che questo fenomeno sia collegato a un indebolimento, e anche a un vero «intasamento» della circolazione atmosferica come lasciano intravedere per esempio i grandi spostamenti della corrente a getto verso sud (nella parte orientale del continente americano) e verso nord (in quella occidentale).

Abbiamo visto che il sistema climatico ha una lunga storia di fenomeni di transizione complessi; indaghiamo perciò su alcuni fattori che possono essere all'origine di questo comportamento complesso. Come nel paragrafo precedente manterremo la discussione sul piano qualitativo; un approccio piú quantitativo verrà dato nel capitolo VI.

Il fattore principale che controlla il clima della Terra è indubbiamente l'afflusso di energia solare. Chiamiamo costante solare Q l'energia solare per unità di tempo e di superficie ricevuta dallo strato superiore dell'atmosfera nella direzione perpendicolare ai raggi del Sole. Calcoli astronomici mostrano che la luminosità del Sole (come la luminosità delle altre stelle della cosiddetta sequenza principale) aumenta nel tempo. Ne segue che alcune centinaia di milioni di anni fa il valore di Q era piú basso di una buona percentuale rispetto al valore attuale. Abbiamo anche osservato che durante l'èra attuale il clima è stato molto mite, e non sono avvenute glaciazioni. Questo «paradosso del Sole

antico debole» mostra che il sistema climatico non segue passivamente la variazione dell'energia solare. In altre parole, i processi generati internamente – soprattutto in questo caso la costituzione dell'atmosfera, le nuvole, e il vigore della circolazione atmosferica – introducono una forte e non trascurabile modulazione degli effetti esterni.

Questa osservazione vale anche per la risposta climatica a un secondo fattore esterno, che nasce dalle variazioni delle caratteristiche orbitali della Terra. I calcoli astronomici mostrano che, a causa dell'azione perturbatrice degli altri corpi del sistema solare, si verificano principalmente tre variazioni dell'orbita terrestre, discusse dalla meccanica elementare. Primo, l'asse di rotazione della Terra precede come una trottola, sicché la direzione celeste verso la quale punta l'asse si sposta attorno a una circonferenza. Secondo, l'angolo di inclinazione dell'asse terrestre con la direzione del Sole (ora di 23,5°) varia nel tempo. Terzo, l'eccentricità dell'orbita terrestre è anch'essa soggetta a variazioni. I periodi caratteristici degli effetti sopra citati sono, rispettivamente, di circa 22 000, 41 000 e 100 000 anni. La teoria prevede che i primi due effetti modifichino la distribuzione stagionale dell'energia solare, mentre il terzo influisca sulla media annuale di energia ricevuta dalla Terra. In tutti i casi gli effetti sono deboli. Come si è detto sopra, la periodicità dominante dei cicli di glaciazione è segnata da questi effetti. Ancora una volta, la dinamica interna dovrebbe giocare un ruolo non trascurabile amplificando questo debole effetto esterno, e innescando cosí un piú grande mutamento climatico.

Diamo ora un esempio di un tale meccanismo di amplificazione. Consideriamo l'interazione fra i processi radiativi e le calotte glaciali. È ben noto che la riflettività del ghiaccio – l'*albedo* – è molto alta. Supponiamo che, a causa di una perturbazione la cui origine non viene qui specificata, una calotta glaciale avanzi leggermente verso l'equatore. A causa dell'aumento dell'albedo globale che ne risulta, la Terra assorbirà una minor energia, e questo fatto aumenterà ulteriormente il raffreddamento del sistema. Ne segue che la calotta glaciale avanzerà ancora, con un conseguente ulteriore abbassamento della temperatura. Siamo dunque in presenza di un ciclo di *feedback* positivo molto simile alla autocatalisi chimica. Si noti che questa è solo una delle numerose e complesse interazioni che operano all'interno del sistema climatico.

Bisognerebbe anche tenere conto del fatto che la Terra riceve la maggior parte dell'energia solare nelle regioni equatoriali e respinge nello spazio buona parte della minore quantità di energia ricevuta dalle regioni polari. In questo modo il flusso solare stabilisce e mantiene un

gradiente di temperatura meridionale fra l'equatore e i poli. Questo è simile al sistema vincolato termicamente incontrato nel § 3 e implica che il sistema climatico viene mantenuto *lontano dall'equilibrio termodinamico*.

Non-equilibrio, *feedback*, fenomeni di transizione, evoluzione... Troviamo nuovamente gli elementi già incontrati nei precedenti esempi di comportamento complesso, e che operano in questo caso nella ben piú impressionante scala planetaria.

9. *Forze contro correlazioni: un riassunto.*

Un pezzo di ferro a una temperatura superiore ai 1044°K non presenta osservabili proprietà magnetiche. Ma se lo raffreddiamo al di sotto di questa temperatura «critica», il materiale diventa magnetizzato. Questo è un tipico esempio di un'importante classe di fenomeni naturali chiamati *transizioni di fase*. Sopra la temperatura critica il materiale è isotropo, nel senso che nessuna delle proprietà osservabili è caratterizzata da una direzione preferenziale. Ma al di sotto di questa temperatura si verifica una magnetizzazione, che è una quantità vettoriale che punta in una certa direzione dello spazio. Il materiale diventa anisotropo, ossia, in termini piú tecnici, si rompe la *simmetria rotazionale* che caratterizza lo stato di non magnetizzazione.

Vi sono molti altri esempi di questo genere. Un liquido è uno stato della materia nel quale le molecole si muovono in tutte le direzioni e non si riconoscono l'una con l'altra lungo distanze maggiori di poche centinaia di milionesimi di centimetro (cfr. figura 1.1*a*). Esso può dunque essere guardato come un materiale omogeneo nel quale tutti i punti dello spazio sono equivalenti. Raffreddiamo ora questo sistema lentamente e uniformemente. Sotto questa temperatura caratteristica (0°C per l'acqua pura alla pressione normale), otteniamo un reticolo cristallino, una fase nuova, solida, della materia. Le sue diverse proprietà, come per esempio la densità, non sono piú uguali se ci muoviamo lungo certe direzioni dello spazio; in altre parole, si rompe la *simmetria traslazionale* che caratterizza il liquido.

In entrambi i casi la rottura di simmetria è concomitante all'apparizione di nuove proprietà che ci fanno ritenere il materiale *ordinato*. Per esempio, nel reticolo cristallino le molecole effettuano delle piccole vibrazioni attorno alle posizioni spaziali regolarmente sistemate, che, a seconda del caso, possono giacere ai vertici di un cubo, ai vertici di un

prisma esagonale regolare, al centro delle loro basi esagonali, e cosí via.

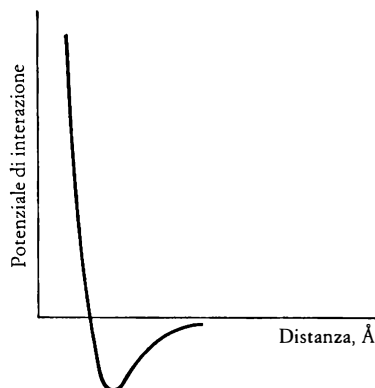
Cosa sta succedendo? Abbiamo semplicemente dimenticato di includere questi fenomeni nella nostra lunga lista di transizioni verso l'autoorganizzazione e il comportamento complesso?

In assenza di interazioni, le molecole di un materiale si muovono liberamente in tutte le direzioni, e questo moto è interamente caratterizzato dalle loro energie cinetiche individuali 1/2 $(m_i v_i^2)$ (m_i, v_i sono le masse molecolari e le velocità), che rimangono invariate nel tempo. Sappiamo tuttavia che in ogni sistema fisico le molecole interagiscono, essenzialmente a causa di forze di origine elettromagnetica. Un tipico esempio sono le forze di Van der Waals che prevalgono in un fluido elettricamente neutro, illustrate nella figura 1.18. Queste sono forze a *corto raggio*, ovvero la loro intensità scende brutalmente verso piccolissimi valori al di là di una separazione interparticellare di pochi diametri molecolari.

Quando il sistema è diluito o quando la temperatura è molto alta, le forze intermolecolari non sono molto efficienti: predomina l'energia cinetica, e il materiale si comporta in modo disordinato. Ma quando abbassiamo la temperatura (o comprimiamo il sistema), i ruoli dell'energia cinetica e delle forze intermolecolari tendono a invertirsi.

Alla fine il sistema viene dominato dalle interazioni e adotta una configurazione nella quale la corrispondente energia potenziale è la piú

Figura 1.18

Una tipica forma del potenziale di interazione intermolecolare in un gas neutro. La distanza che corrisponde al minimo della curva è di pochi angstrom, e il valore dell'energia potenziale a questo punto è circa $k_B T_c$, *dove* k_B è la costante di Boltzmann e T_c la temperatura critica della transizione liquido-vapore.

bassa possibile. Essendo interamente mediata dalle interazioni intermolecolari, essa mostra una lunghezza caratteristica – come la zona reticolare – che è microscopica (pochi angstrom), confrontabili al raggio di azione di queste interazioni. Inoltre, il comportamento del materiale è indipendente dal tempo se l'ambiente stesso è stazionario. In questo senso si può dire che l'ordine associato a una transizione di fase porta a oggetti «fossili».

Entrambe le proprietà debbono essere contrapposte a quelle che caratterizzano i fenomeni di transizione illustrati nei precedenti paragrafi di questo capitolo. Infatti, le dimensioni spaziali che caratterizzano gli stati che emergono dopo queste ultime transizioni sono *macroscopiche*, paragonabili o piú grandi delle dimensioni caratteristiche dei sistemi biologici. Inoltre, in aggiunta alle strutture spaziali, può nascere e venir mantenuta una varietà di fenomeni dipendenti dal tempo, le cui dimensioni caratteristiche sono nuovamente macroscopiche.

La ragione che sta dietro a questa differenza è che agisce un nuovo fattore di organizzazione non riconducibile a interazioni intermolecolari: il *vincolo di non-equilibrio*. Poiché include parti del sistema che contengono un gran numero di molecole, scaturiscono nuovi fenomeni associati a stati della materia che godono di correlazioni a lungo raggio. È istruttivo comprendere come in una struttura dendritica come un fiocco di neve questi due tipi di ordine sono sovrapposti e possono essere facilmente distinti e contrapposti. Invero il reticolo cristallino ha poco a che fare con il formarsi delle dendriti, le loro dimensioni e la loro distanza, che sono maggiori di parecchi ordini di grandezza di qualsiasi lunghezza cristallografica. In breve, il comportamento complesso a cui siamo interessati può essere pensato come una transizione di fase di un nuovo tipo, nella quale l'abbassamento della temperatura viene sostituito dall'applicazione e dal progressivo aumento di adatti vincoli di non-equilibrio. Per questa ragione il termine «fase di transizione di non-equilibrio» verrà usato talvolta in connessione con i fenomeni autoorganizzati che portano al comportamento complesso.

I fenomeni di rottura di simmetria mediati da forze, simili sotto alcuni aspetti alle transizioni di fase, appaiono anche a livello delle interazioni fondamentali. Uno dei piú sorprendenti esempi è la rottura di simmetria fra materia e antimateria. Come è ben noto, a ogni particella carica la fisica quantistica relativistica associa un'antiparticella della stessa massa, ma di carica opposta: abbiamo elettroni e antielettroni (positroni), protoni e antiprotoni, ecc. Le antiparticelle sono state osservate sperimentalmente in condizioni speciali. Tuttavia, sebbene le

equazioni fondamentali della fisica siano simmetriche rispetto alla materia e all'antimateria, la proporzione di antimateria nell'universo osservabile è trascurabile rispetto alla materia. Da dove viene questa asimmetria? Una possibile risposta viene dalle teorie unificate delle interazioni forte, elettromagnetica, e debole delle particelle elementari, che si ritiene descrivano il comportamento della materia a temperature molto alte, quelle, per esempio, che prevalsero subito dopo l'inizio del nostro universo. In particolare, sembra che la materia sia passata attraverso una o forse piú transizioni alle «temperature critiche» quando venne raffreddata, a causa dell'espansione dell'universo, dallo stato iniziale estremamente caldo. Secondo le grandi teorie unificate, la materia sopra la prima temperatura di transizione era in uno stato fortemente simmetrico. I quark, i costituenti dei protoni, dei neutroni e dei leptoni, che includono gli elettroni, i neutrini e le loro antiparticelle, si comportavano tutti in modo identico. Al di sotto del punto di transizione, tuttavia, le differenze si manifestarono attraverso la rottura di simmetria. Alla fine queste particelle differenziate divennero il materiale grezzo che costituisce le stelle, i pianeti e gli esseri viventi. In questo senso, dunque, una transizione mediata dalle sopra citate tre interazioni fondamentali a corto raggio ha raggiunto una scala macroscopica che comprende l'attuale universo, che sembra essere il «relitto» di un evento prodotto nel lontano passato. Non c'è contraddizione, tuttavia, fra questa affermazione e il nostro sforzo precedente di contrapporre le normali transizioni di fase alle transizioni di fase di non-equilibrio sulla base dei loro raggi d'azione molto diversi. La storia dell'inizio dell'universo e, in particolare, la sua espansione, è effettivamente segnata dal passaggio da uno stato di equilibrio termico a uno stato nel quale viene rotto sia l'equilibrio fra diversi costituenti della materia sia quello fra materia e radiazione. In un tale ambiente di non-equilibrio possono avere luogo transizioni di rottura di simmetria su larga scala come quelle menzionate piú sopra. Sotto questo aspetto dunque, la materia differenziata come la osserviamo oggi si può pensare come il risultato di un primordiale non-equilibrio.

La scoperta di questi processi di non-equilibrio primordiale ha profonde conseguenze sul nostro modo di vedere l'evoluzione dell'universo. Nel capitolo II, affronteremo l'universalmente nota seconda legge della termodinamica, che tratta dell'aumento dell'entropia. Può essere interessante ricordare che, nella sua formulazione originale, Rudolf Clausius enunciò questo principio in termini cosmologici: «L'energia dell'universo è costante; l'entropia dell'universo cresce costantemente

verso un valore massimo». Cosí l'evoluzione termodinamica dell'universo andrebbe da un livello basso di entropia a uno alto. Questo portò all'idea della «morte termica» che dovrebbe essere il nostro inesorabile destino. Recenti sviluppi della cosmologia stanno tuttavia mettendo in discussione questo modo di vedere. Nel 1965, esattamente un secolo dopo la formulazione della seconda legge da parte di Clausius, Penzias e Wilson scoprirono la radiazione fossile, mostrando che l'universo osservabile è ricco di fotoni. Di fatto esiste un solo barione ogni 10^8 o 10^9 fotoni. La materia essenzialmente galleggia in un mare di fotoni. Questo significa che il fattore piú importante dell'entropia è di gran lunga costituito da questi fotoni, che sono stati prodotti durante le fasi iniziali dell'universo.

Per descrivere il processo irreversibile che ha portato dalla situazione iniziale a un universo riccamente differenziato e strutturato è necessario tenere conto della gravità, la quarta forza fondamentale della natura, ben nota per il suo carattere di azione a lungo raggio. La teoria moderna della gravitazione è stata sviluppata sulla scorta alla teoria della relatività generale di Einstein. Vediamo di dare un'idea dei rapporti che possono esistere tra gravitazione, relatività generale e processi irreversibili, rinviando all'*Appendice* E per maggiori dettagli.

Come è noto, il mondo, nella descrizione data dalla teoria newtoniana presenta un'intrinseca dualità: da una parte lo spazio-tempo, dall'altra la materia soggetta all'influenza delle forze. La grande idea di Einstein fu di superare questa dualità e di mettere lo spazio-tempo in relazione con la materia. La relatività generale è la base di tutti i modelli di evoluzione cosmologica.

Vi sono tuttavia alcune difficoltà. Le equazioni di Einstein descrivono un'evoluzione in cui materia ed entropia si conservano; pertanto, se andiamo indietro fino al passato piú lontano, arriviamo a una singolarità in cui tutta la materia e l'entropia sono concentrate in un punto: questa è la famosa singolarità del Big Bang. A. Wheeler ha definito questa singolarità «la piú grande crisi della fisica» e in effetti al di là di essa le leggi fisiche non hanno piú significato. Oggi molti tentativi vengono fatti per evitare questa singolarità e sostituirla con una qualche fase di transizione. Ciò che vogliamo mettere qui in rilievo è che, durante questa fase primordiale, i processi irreversibili hanno probabilmente giocato un ruolo essenziale.

Di fatto, nella nostra visione delle cose, la stessa apparizione della materia dal vuoto costituirebbe il processo irreversibile fondamentale dell'inizio dell'universo.

Capitolo secondo

Il vocabolario della complessità

Nel precedente capitolo abbiamo adottato il punto di vista naturalistico: ci siamo guardati attorno e abbiamo cercato di formarci un'idea intuitiva della complessità osservando i fenomeni di transizione che portano all'autoorganizzazione in diversi contesti. Cosí facendo, abbiamo incontrato un certo numero di nuove nozioni, che ricorrevano in continuazione nei nostri esempi. Ora è tempo di esaminare queste nozioni in modo piú sistematico e approfondito e costruire cosí il *vocabolario della complessità*. Non pretendiamo che alla fine del capitolo il lettore abbia gli strumenti per usare questi concetti in modo creativo: questo sarà l'obiettivo dei capitoli III, IV e V.

1. *Sistemi conservativi.*

Fin dalla nascita del pensiero astratto che segnò l'inizio della scienza e della moderna civiltà, il concetto di *conservazione* giocò un ruolo chiave. Il grande pensatore Talete, considerato tradizionalmente come il primo dei filosofi, aveva la convinzione che le molte sfaccettature della natura potevano essere ridotte a una singola sostanza fondamentale, l'acqua. In termini moderni, egli vedeva l'universo come un *sistema conservativo*, nel quale, nonostante il continuo scambio che avviene fra le sue diverse parti, esisteva un elemento primordiale che rimaneva intoccabile sotto l'effetto di ogni mutamento. Simili pensieri devono aver influenzato molto il giovane Platone, la cui teoria delle idee è essenzialmente una brillante ricerca di stabilità a un livello piú astratto.

Duemila anni dopo incontriamo un altro grande periodo della storia umana, quello di Newton e Leibniz. Con l'avvento della meccanica classica, l'idea di conservazione venne incredibilmente allargata: la conservazione non poggiava piú su qualche incontrollabile assioma, ma, piuttosto, era la conseguenza di una serie di leggi che fornivano la spiegazione quantitativa delle osservazioni sul moto dei corpi celesti e sul

moto degli oggetti sotto l'effetto della gravità terrestre. Invero, la seconda legge di Newton (che mette in relazione la forza con l'accelerazione) e la terza legge (che mette in relazione l'«azione» con la «reazione»), implicano che, in un sistema di punti materiali interagenti, liberi da ogni forza esterna, le seguenti tre quantità rimangono invarianti nel tempo:

1. Energia totale
2. Il momento traslazionale totale
3. Il momento angolare totale

1. L'energia totale,

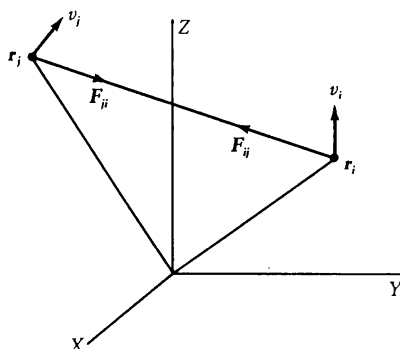
$$E = \text{energia cinetica} + \text{energia potenziale}$$

$$= \sum_i \frac{1}{2} m_i v_i^2 + \frac{1}{2} \sum_{i \neq j} V_{ij} \qquad (2.1a)$$

in cui m_i e $\boldsymbol{v}_i$ sono rispettivamente la massa e la velocità del punto materiale i, e V_{ij} è il potenziale di interazione fra i e j, legato alla corrispondente forza $\boldsymbol{F}_{ij}$ attraverso la relazione:

lavoro della forza = cambio dell'energia potenziale

Figura 2.1

Forze centrali di interazione esercitate fra due particelle poste in $\boldsymbol{r}_i$, $\boldsymbol{r}_j$ e aventi velocità $\boldsymbol{v}_i$, $\boldsymbol{v}_j$. Dalla terza legge di Newton, $\boldsymbol{F}_{ij} = -\boldsymbol{F}_{ji}$.

Per una forza che dipende solo dalla distanza $\boldsymbol{r} \equiv \boldsymbol{r}_i - \boldsymbol{r}_j$, questa si riduce a

$$F_{ij}(\boldsymbol{r}) = -\nabla_r V_{ij}(\boldsymbol{r}) \tag{2.1b}$$

dove ∇_r è la derivata del potenziale rispetto alla distanza.

2. Il momento traslazionale totale,

$$\boldsymbol{P} = \sum_i \boldsymbol{p}_i = \sum_i m_i \boldsymbol{v}_i \tag{2.2}$$

3. Il momento angolare totale,

$$\boldsymbol{L} = \sum_i \boldsymbol{l}_i = \sum_i m_i \boldsymbol{r}_i \times \boldsymbol{v}_i \tag{2.3}$$

nel quale $\boldsymbol{l}_i$ descrive la parte rotazionale del moto di ogni punto materiale attorno a un certo punto dello spazio scelto per comodità nell'origine del sistema di coordinate.

Tutte queste leggi di conservazione sono intimamente legate a una proprietà chiave delle equazioni base della meccanica, e cioè che la quantità dinamicamente collegata alla forza (generalmente una funzione delle coordinate $\boldsymbol{r}_i$) è l'accelerazione, la derivata seconda dello spazio (e non la velocità, come pensava Aristotele!):

$$m_i \frac{d^2 \boldsymbol{r}_i}{dt^2} = \sum_j F_{ij}(\boldsymbol{r}_i - \boldsymbol{r}_j) \tag{2.4}$$

Un osservatore, dunque, che indaghi il moto nel passato ($t' = -t$) noterà esattamente le stesse proprietà del normale osservatore terrestre, che compie misure per tempi «positivi». In altre parole, osservando che $d^2\boldsymbol{r}/dt^2 = d\boldsymbol{v}/dt$ e che $\boldsymbol{v} = d\boldsymbol{r}/dt$, si vede che la trasformazione $t' = -t$ induce l'ulteriore trasformazione $\boldsymbol{v}' = -\boldsymbol{v}$, lasciando invariata l'accelerazione: i punti materiali vengono mantenuti lungo le stesse traiettorie, ma la direzione che essi seguono viene invertita. Non c'è nessuna particolarità in un cambiamento del genere, poiché, ad esempio, avremmo benissimo potuto comunicare loro delle velocità tali che questa direzione venisse seguita per tempi positivi. In altre parole, niente nelle equazioni del moto ci permette di distinguere il moto diretto da quello inverso.

La meccanica classica appare cosí la scienza dei sistemi conservativi e invertibili rispetto al tempo. Un'elegante trattazione che mostra que-

sti aspetti è basata sulla formulazione di Hamilton, nella quale la struttura delle equazioni del moto è direttamente collegata all'energia – la piú importante delle tre quantità che si conservano. In particolare, introducendo la funzione *hamiltoniana*, H

$$H = H(p, q) = E = \text{costante}$$

in cui p, q generalizzano il momento ($p = mv$) e la posizione ($q = r$) della formulazione newtoniana, si ha:

$$H = \sum_i \frac{p_i^2}{2m_i} + \frac{1}{2} \sum_{ij} V_{ij}(q_i - q_j) \tag{2.5a}$$

Si vede subito che l'equazione 2.4, come pure la relazione $v = dr/dt$ sono equivalenti alle:

$$\begin{aligned} \frac{d}{dt} q_i &= \frac{\partial H}{\partial p_i} \\ \frac{d}{dt} p_i &= -\frac{\partial H}{\partial q_i} \end{aligned} \tag{2.5b}$$

Oggi la dinamica hamiltoniana è un'importante parte della scienza. Nel capitolo III descriveremo brevemente i rivoluzionari sviluppi che la caratterizzarono fin dagli anni '50. Chiudiamo ora questo paragrafo dando alcuni esempi di sistemi conservativi.

Uno dei piú familiari sistemi conservativi è il pendolo (vedi figura 1.8). Se indichiamo con l la lunghezza, con g l'accelerazione di gravità, con θ l'angolo misurato dalla verticale, e decidiamo di misurare le energie usando come riferimento la posizione della massa sulla verticale, possiamo scrivere l'energia totale – l'hamiltoniana – nella forma

$$\begin{aligned} H &= \text{energia cinetica} + \text{energia potenziale} \\ &= \frac{1}{2} mv^2 + mgl(1 - \cos\theta) \end{aligned}$$

Poiché v è tangente alla traiettoria della massa, $v = l(d\theta/dt)$, e quindi

$$H = ml\left[\frac{1}{2}\, l\left(\frac{d\theta}{dt}\right)^2 + g(1 - \cos\theta)\right] = \text{costante} \tag{2.6a}$$

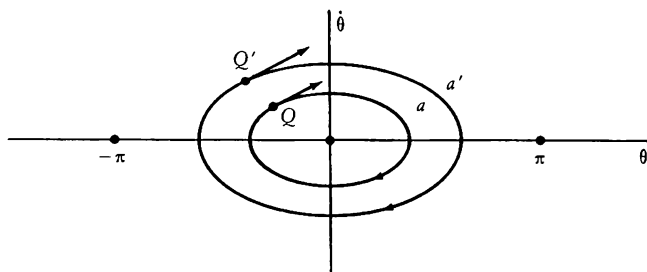
Come verrà discusso dettagliatamente nel capitolo III, è molto utile visualizzare l'evoluzione di un sistema in uno spazio astratto, noto co-

me lo *spazio delle fasi*, nel quale le coordinate sono le posizioni e i momenti che appaiono nell'equazione di Hamilton (cfr. le equazioni 2.5*b*). Nel caso del pendolo, $q = \theta$, $p = ml(d\theta/dt) \equiv ml\dot{\theta}$, lo spazio delle fasi è il piano $(\dot{\theta}, \theta)$. Lo stato istantaneo del pendolo è dunque un punto in questo spazio, chiamato *punto rappresentativo*. Al variare del tempo sia θ che $\dot{\theta}$ assumono un unico e ben definito valore; in altre parole il punto rappresentativo descrive una curva unidimensionale – la traiettoria nello spazio delle fasi. La tangente a questa curva sarà la velocità nello spazio delle fasi (cfr. figura 2.2). Queste non debbono essere confuse con i piú familiari termini traiettoria e velocità di un corpo che si muove in uno spazio fisico, i quali, nonostante la loro utilità, forniscono una descrizione meno dettagliata del moto.

Un caso particolare della traiettoria nello spazio delle fasi è la posizione di equilibrio, nella quale la velocità e l'accelerazione sono nulli. Questa è una traiettoria degenere, poiché è rappresentata da un unico punto nello spazio delle fasi. In un pendolo vi sono due di questi punti di equilibrio (figura 2.2), che giacciono lungo la verticale, in basso $(\theta = 0)$ e in alto $(\theta = \pm\pi)$. Come vedremo tra breve, essi posseggono uno status completamente diverso.

Abbiamo appreso dall'osservazione e dai corsi elementari di meccanica che a seguito di piccole velocità iniziali e deviazioni dalla posizione inferiore della verticale, il pendolo compie oscillazioni prolungate nel tempo sotto forma di piccole vibrazioni. Questo significa che sia θ che $\dot{\theta}$ assumono gli stessi valori dopo ogni periodo. Nella rappresentazione dello spazio delle fasi, questo viene descritto da una traiettoria che è

Figura 2.2

Lo spazio delle fasi $(\dot{\theta}, \theta)$ del pendolo $(-\pi \leqslant \theta \leqslant \pi)$. $\theta = \dot{\theta} = 0$ equilibrio stabile; $\dot{\theta} = 0$, $\theta = \pm\pi$: equilibrio instabile; a, a': traiettorie periodiche corrispondenti a due diversi valori di energia; Q, Q': punti rappresentativi su queste traiettorie.

una *curva chiusa* (*a* oppure *a'* nella figura 2.2), lungo la quale l'energia, equazione 2.6, viene conservata.

Un altro sistema conservativo familiare è l'oscillatore armonico, che rappresenta il moto senza attrito di una particella di massa *m* appesa a una corda orizzontale, perfettamente elastica, il cui secondo estremo è rigidamente fissato. La forza esercitata sulla particella è proporzionale alla deviazione *x* della lunghezza della corda rispetto al suo valore all'equilibrio. Denotiamo con *k* il coefficiente di proporzionalità e scriviamo l'hamiltoniana del sistema come la somma dell'energia cinetica della particella e dell'energia elastica della molla:

$$H = \frac{1}{2}m\dot{x}^2 + \frac{1}{2}kx^2 \qquad (2.6b)$$

Una ancora diversa classe di sistemi conservativi nasce dal problema dei due corpi, ovvero il moto di due corpi interagenti a causa della gravitazione o di una qualsiasi altra forza dipendente solo dal valore assoluto della loro distanza reciproca. Le complicazioni nascono tuttavia facilmente in meccanica. Presi due pendoli accoppiati per mezzo di una molla, o un pendolo periodicamente forzato dall'esterno, o tre masse interagenti (il celebre problema dei tre corpi), arriviamo al confine della conoscenza attuale. Discuteremo in dettaglio questi problemi nel capitolo III.

Non si deve concludere dalla precedente discussione che i sistemi conservativi sono confinati alla meccanica. La propagazione della luce nel vuoto, descritta dalle celebri equazioni di Maxwell, ci fornisce un importante esempio di sistemi conservativi nel contesto dell'elettromagnetismo. Sistemi conservativi si hanno anche in problemi di meccanica quantistica che trattano le proprietà degli atomi e delle molecole in assenza di campi esterni.

2. *Sistemi dissipativi.*

In aggiunta ai sistemi conservativi descritti dalla dinamica classica, dobbiamo considerare i sistemi che dànno luogo a processi irreversibili. Un esempio molto semplice è fornito dai sistemi che presentano attrito.

L'importanza dell'attrito, una particolare forma di dissipazione, venne notata molto prima che la dinamica classica venisse formulata. Quando Aristotele introdusse l'ipotesi che i sistemi dinamici sublunari tendevano in generale a uno stato di riposo, esprimeva l'idea che un

qualche tipo di «attrito» rallentasse il moto. Sotto questo aspetto il classico principio di inerzia, che afferma che la quantità meccanica fondamentale è l'accelerazione, e non la velocità, corrisponde a una idealizzazione nella quale si trascura l'attrito.

A partire dal lavoro di Fourier e Clausius, l'interesse per i sistemi dissipativi che dànno luogo a processi irreversibili divenne sempre piú diffuso nel XIX secolo. Ciò è del tutto naturale in un periodo che ha assistito alla rivoluzione industriale. In questo contesto, tuttavia, la dissipazione sembrava essere associata alla degradazione, all'esaurirsi della energia disponibile.

È interessante notare come fra i filosofi antichi, Platone fosse ben conscio che sia la stabilità sia il cambiamento debbano essere parti integranti della realtà. Ma nel XIX secolo si determinò una situazione conflittuale. L'irreversibilità e la dissipazione erano viste, in fisica, come degradazione; mentre l'evoluzione biologica – che ovviamente è un processo irreversibile – veniva associata a un aumento di complessità. A causa delle sue implicazioni tecnologiche, la meccanica dei fluidi fu il primo campo nella storia della scienza in cui i processi dissipativi giocarono un ruolo cruciale. Ma non appena l'idea della costituzione molecolare della materia si impose gradualmente nella scienza, la cinetica chimica, il moto browniano e vari tipi di fenomeni di trasporto seguirono questa tendenza. Oggi si è compreso che i sistemi dissipativi costituiscono una grande e importante classe di sistemi naturali. Molti degli esempi del capitolo I trattavano infatti fenomeni che hanno luogo in sistemi come questi, non ultima la vita stessa.

La piú importante differenza fra sistemi conservativi e dissipativi appare quando si tenta di fornire una *descrizione macroscopica* di questi ultimi, in cui le variabili globali come la temperatura, la concentrazione, la pressione, la convezione, la velocità, ecc. vengono usate per definire lo stato istantaneo. Quando si studiano le equazioni di evoluzione di queste variabili, ne appare un aspetto generale, piú precisamente che esse non sono invarianti per inversione temporale, contrariamente alle equazioni 2.4 e 2.5. Su queste basi ci si aspetta dunque un corso irreversibile degli eventi.

Un'illustrazione sorprendente di questa situazione è data da una reazione chimica. Consideriamo, per esempio, il processo descritto dall'equazione 1.2*a*. La velocità con cui le particelle della specie A si consumano è proporzionale alla frequenza di incontri con le molecole A e B, frequenza che, se il sistema è diluito, è proporzionale unicamente alla concentrazione.

Cosí

(2.7*a*) $$\frac{dc_A}{dt} = -kc_Ac_B$$

Evidentemente, se invertiamo il tempo, $t' = -t$, e se chiamiamo c_A', c_B' i valori delle concentrazioni in funzione di t', l'equazione 2.7 diventa

$$\frac{dc_A'}{dt'} = kc_A'c_B'$$

e descrive un processo nel quale c_A viene prodotta invece di essere consumata. Questo non è certamente equivalente al fenomeno descritto dall'equazione 2.7*a*.

Altri esempi di processi dissipativi sono la conduzione di calore o la diffusione. Gli esperimenti mostrano che quando una leggera disomogeneità appare entro un fluido uniforme, alla fine si allarga e scompare. La stessa evoluzione a senso unico viene osservata quando una leggera variazione di temperatura viene applicata localmente e momentaneamente a un fluido isotermico. Le equazioni seguenti, conosciute rispettivamente come le equazioni di Fick e di Fourier, forniscono una descrizione quantitativa di questi fenomeni, che è in perfetto accordo con le osservazioni sperimentali:

(2.7*b*) $$\frac{\partial c}{\partial t} = D\nabla^2 c \qquad D > 0$$ (equazione di Fick)

(2.7*c*) $$\frac{\partial T}{\partial t} = \kappa\nabla^2 T \qquad \kappa > 0$$ (equazione di Fourier)

dove c denota la concentrazione di una certa sostanza disciolta nel fluido, T la temperatura, D e κ il coefficiente di diffusione della massa e del calore rispettivamente. Di nuovo, se invertiamo il tempo otteniamo le leggi completamente diverse

$$\frac{\partial c'}{\partial t'} = -D\nabla^2 c'$$

$$\frac{\partial T'}{\partial t'} = -\kappa\nabla^2 T'$$

che descrivono una situazione in cui una concentrazione iniziale o una perturbazione termica verrebbero amplificate anziché smorzate.

Le variabili concentrazione e temperatura sono esempi di *variabili pari*, i cui segni non cambiano con l'inversione temporale. In contrapposizione, il momento di una particella o la velocità di convezione di un fluido sono *variabili dispari*, poiché si esprimono, in ultima analisi, come le derivate rispetto al tempo di variabili come la posizione e cambiano il segno con l'inversione temporale. Questo porta alla seguente proprietà generale dell'equazione di evoluzione di un sistema dissipativo. Siano $X_1, \dots, X_n$ l'insieme delle variabili macroscopiche di un tale sistema. La loro evoluzione nel tempo è:

$$\frac{\partial X_i}{\partial t} = F_i(X_1, \dots, X_n, \boldsymbol{r}, t, \dots) \tag{2.8}$$

in cui F_i possono essere funzioni complicate delle X e delle loro derivate spaziali, come anche funzioni esplicite dello spazio $\boldsymbol{r}$ e del tempo t. Allora, in un sistema dissipativo, se eseguiamo un'inversione del tempo, $t' = -t$, almeno una delle funzioni derivate F_i corrispondente a una variabile pari X_i deve contenere una parte che rimane invariante sotto inversione temporale, mentre la funzione derivata F_i che corrisponde a una variabile dispari X_i deve contenere una parte che cambia segno sotto inversione temporale. Un esempio di derivate del primo tipo è dato dal membro a destra delle equazioni da 2.7*a* a 2.7*c*, mentre un esempio del secondo tipo è dato dal contributo della viscosità all'equazione dell'equilibrio del momento in un fluido soggetto a convezione.

Un cruciale passo in avanti nella distinzione fra sistemi conservativi e dissipativi verrà fatto nel § 5, quando sarà discussa la seconda legge della termodinamica e le sue implicazioni.

Come nei sistemi conservativi, anche nei sistemi dissipativi si può introdurre un conveniente spazio delle fasi. È generato dall'insieme delle variabili presenti e diventa cosí, nel caso di un mezzo continuo nel quale le varie proprietà vengono distribuite nello spazio, uno spazio di un infinito numero di dimensioni (cfr. le equazioni 2.7*b*, 2.7*c*). Lo spazio delle fasi è molto piú facile da trattare nel caso in cui il numero delle variabili è discreto e, in particolare, finito o preferibilmente piccolo. Nell'esempio dell'equazione 2.7*a*, lo spazio delle fasi si riduce a una linea, e la traiettoria segue necessariamente questa linea (figura 2.3*a*). Un esempio meno banale è dato dallo schema

$$A \xrightarrow{k_1} X \xrightarrow{k_2} Y \xrightarrow{k_3} F \tag{2.9a}$$

con

$$\frac{dc_x}{dt} = k_1c_A - k_2c_X$$
(2.9b)
$$\frac{dc_y}{dt} = k_2c_X - k_3c_Y$$

Le traiettorie nello spazio delle fasi sono mostrate nella figura 2.3*b*.

È istruttivo capire che certi sistemi dissipativi possono essere trasformati in conservativi, ed essere descritti secondo la schema hamiltoniano. Un celebre esempio è il meccanismo Lotka-Volterra

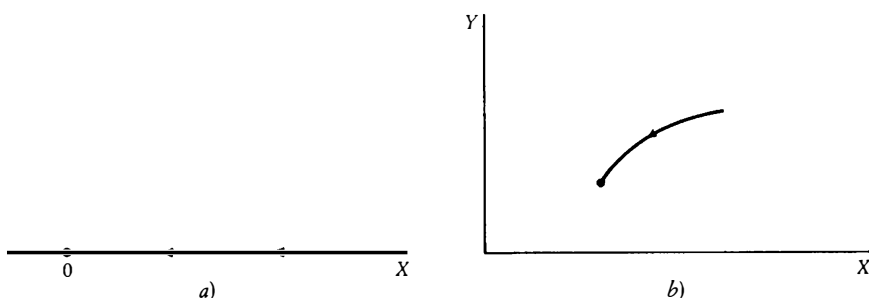
$$A + X \xrightarrow{k_1} 2X$$
(2.10)
$$X + Y \xrightarrow{k_2} 2Y$$
$$Y \xrightarrow{k_3} D$$

nel quale esiste una costante non banale del moto, che gioca il ruolo dell'hamiltoniana. Tuttavia, nonostante il carattere apparentemente conservativo, l'inversione del tempo non vale poiché sia X che Y sono variabili positive. È dunque privo di significato attribuire loro proprietà simili a quelle del momento in meccanica classica, proprietà necessarie per assicurare una tale invarianza.

Una questione basilare qui non discussa, è come i sistemi dissipativi

Figura 2.3

Rappresentazione dello spazio delle fasi di un sistema dissipativo.

a) Sistema a una variabile dell'equazione 2.7*a*.
b) Sistema a due variabili delle equazioni 2.9*b*.

siano correlati con quelli conservativi, e se si possa trovare il passaggio da una descrizione all'altra. Dedicheremo a questo problema buona parte del capitolo v.

3. *Equilibrio meccanico e termodinamico. Vincoli di non-equilibrio.*

In meccanica, l'equilibrio è uno stato particolare nel quale sia le velocità che le accelerazioni di tutti i punti materiali sono uguali a zero. Abbiamo già incontrato questo tipo di «stato di riposo» nel capitolo I, §§ 1 e 4, e nel § 1 di questo capitolo. Per definizione la forza risultante che agisce su ciascun corpo è uguale a zero in ogni istante, altrimenti l'equilibrio verrebbe rotto. Questo è ciò che accade, per esempio, quando si frattura un pezzo di materiale per effetto di un peso.

Alcuni sistemi meccanici familiari funzionano in modo tale che l'equilibrio non si realizza mai. Per esempio, l'orbita della Terra attorno al Sole è una manifestazione di una deviazione permanente dall'equilibrio meccanico. Quest'ultimo può rendersi possibile solo con una separazione infinita fra i due corpi.

La nozione di *equilibrio termodinamico*, ripetutamente incontrata nel capitolo I, §§ 3, 4, 7 e 8 è nettamente differente. Contrariamente all'equilibrio meccanico, le molecole costituenti il sistema sono soggette a forze e si muovono in continuazione in tutte le possibili direzioni a meno che la temperatura non diventi molto bassa. «Equilibrio» è relativo in questo caso a una qualche proprietà collettiva che caratterizza il sistema, come la temperatura, la pressione, o la concentrazione di un costituente chimico. Schematizziamo una formulazione generale di questo concetto, del quale gli esempi esposti nel capitolo I sono casi particolari.

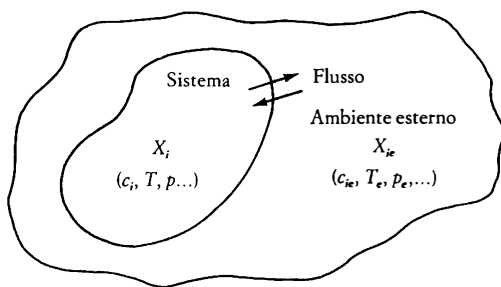
Consideriamo un sistema situato in un certo ambiente (figura 2.4). Conveniamo di caratterizzarli entrambi attraverso un insieme di proprietà, siano X_i e X_{ie}, il cui valore operativo risiede primariamente nella loro relazione con gli *scambi* che avvengono fra il sistema e l'ambiente. Per esempio, se il sistema è contenuto in un recipiente le cui pareti sono perfettamente rigide, permeabili al calore e impermeabili alla materia, X sarà uguale alla *temperatura* T e controllerà lo scambio di energia sotto forma di calore fra il sistema e il suo ambiente.

Diciamo che il sistema è in equilibrio termodinamico se è completamente identificato al suo ambiente, ossia se le proprietà che costituiscono gli insiemi X_i e X_{ie}, hanno valori identici. Nell'esempio precedente, l'equilibrio termodinamico è equivalente a $T = T_e$ in ogni istante e in

ogni punto dello spazio. Tuttavia, poiché le pareti del recipiente sono impermeabili alla materia, il sistema e l'ambiente rimarranno altamente diversificati nella loro composizione chimica c. Se, d'altra parte, le pareti diventano permeabili a certe sostanze chimiche i, l'equilibrio termodinamico prevarrà quando il sistema e l'ambiente diverranno indistinguibili per quanto riguarda le suddette sostanze chimiche. In semplici casi questo significa che le corrispondenti variabili di composizione soddisferanno l'uguaglianza $c_i = c_{ie}$, ma in situazioni piú generali l'equilibrio sarà caratterizzato dall'uguaglianza di una proprietà detta *potenziale chimico*, $\mu_i = \mu_{ie}$. Analogamente, se le pareti del recipiente non sono rigide, il sistema può scambiare energia meccanica con l'ambiente. Allora l'equilibrio implica anche l'uguaglianza della *pressione*, $p = p_e$. Altre variabili che caratterizzano l'equilibrio termodinamico possono nascere in presenza di forze esterne, come, per esempio, un campo elettrico applicato a un mezzo polarizzato, una sollecitazione applicata a un materiale deformabile, come un solido plastico, e cosí via.

Il caso dei *sistemi isolati*, che non possono scambiare né energia né materia con l'ambiente, merita particolare attenzione. Per definizione, non possiamo dire che all'equilibrio un sistema isolato venga identificato con il suo ambiente. Tuttavia, vale buona parte della discussione precedente, poiché una piccola parte di un sistema isolato è soggetto a un ambiente costituito dalle rimanenti parti. Come risultato, all'equilibrio le diverse parti del sistema diverranno indistinguibili.

Secondo le definizioni di cui sopra, l'equilibrio è automaticamente uno stato stazionario, $\partial X_i / \partial t = 0$. È tuttavia uno stato stazionario di tipo particolare, poiché, essendo identica la proprietà X_i in questione, il

Figura 2.4
Rappresentazione schematica di un sistema aperto.

sistema e l'ambiente non hanno nulla da scambiare sistematicamente. Ad esempio, in un equilibrio condizionato dall'uguaglianza della temperatura, della composizione chimica e delle pressioni, non ci sarà scambio sistematico né di energia termica o meccanica, né di materia. Esprimiamo questo dicendo che non ci sono *flussi* netti attraverso il sistema.

$$J_i^{eq} = 0 \tag{2.11}$$

Questa uguaglianza non implica tuttavia che l'equilibrio sia uno stato di completa rigidità. Dovrebbe essere piuttosto inteso in senso dinamico, come spiegato nel capitolo 1, § 4, dedicato alle reazioni chimiche. In particolare, per ogni processo che induce una variazione di J_i (sia per esempio un leggero aumento della temperatura T rispetto a T_e in una piccola parte del sistema), vi sarà un processo inverso che contribuisce alla variazione di J_i nella direzione esattamente opposta. Questa proprietà fondamentale del *bilancio dettagliato* non è altro che una manifestazione della reversibilità nel tempo dei processi elementari che si attuano nel sistema e che obbediscono alle leggi dei sistemi conservativi viste nel § 1. Essa è responsabile della maggior parte delle forti proprietà dell'equilibrio, come ad esempio l'unicità e la stabilità che verranno discusse piú in profondità nei §§ 4-6.

È ora semplice estendere la nostra discussione agli *stati di non-equilibrio*: questi stati verranno associati a flussi non nulli fra il sistema e l'ambiente e, contemporaneamente, a differenze fra le variabili di stato X_i e X_{ie}. Essi possono essere *transienti*, nel senso che possono sorgere momentaneamente a causa di qualche condizione iniziale e scomparire infine quando il sistema si equilibria con il suo ambiente. Ma possono anche essere permanenti, se si stabiliscono e si mantengono determinate condizioni, che chiameremo *vincoli*. Cosí una differenza sistematica di temperatura applicata fra due sezioni di una piastra, o il pompaggio di un prodotto iniziale in un sistema che coinvolge reazioni chimiche accompagnato dalla simultanea evacuazione del prodotto finale, darà luogo a situazioni di non-equilibrio per le quali il sistema non riesce mai a identificarsi con l'ambiente. Questi esempi non devono naturalmente far pensare che il non-equilibrio sia una condizione imposta artificialmente attraverso un esperimento intelligente. Come abbiamo già notato nel capitolo 1, § 8, assistiamo spesso nel nostro ambiente a stati di non-equilibrio, come, per esempio, lo stato della biosfera che è soggetta a un flusso di energia generato dall'equilibrio della radiazione fra Sole e Terra.

A causa dell'azione continua o temporanea del vincolo, in uno stato di non-equilibrio non sussiste il bilancio dettagliato. Come risultato, un regime di non-equilibrio diventa suscettibile di cambiamenti: piccoli e localizzati tentativi per deviare da esso non vengono necessariamente fatti fallire da una contro azione istantaneamente sviluppata, ma piuttosto vengono accettati o anche amplificati dal sistema, diventando cosí sorgenti di innovazione e diversificazione. Come vedremo in seguito, questo è all'origine della capacità dei sistemi al non-equilibrio di essere soggetti a biforcazioni verso nuovi stati che non posseggono alcuna somiglianza con lo stato di equilibrio.

4. *Non-linearità e «feedback».*

Nel paragrafo precedente abbiamo dato una definizione «operativa» degli stati di equilibrio e dei vincoli del non-equilibrio. Adottiamo ora un punto di vista dinamico che punta a stabilire una relazione fra questi concetti e le leggi che governano l'evoluzione delle variabili di stato. Per l'equilibrio meccanico abbiamo infatti già stabilito questa connessione con la discussione fatta alla fine del § 1. In questo paragrafo concentriamo dunque l'attenzione sui sistemi dissipativi, sebbene molti dei concetti che saranno introdotti possono essere tranquillamente trasferiti ai sistemi conservativi.

L'espressione matematica che descrive come evolve nel tempo un insieme di variabili di stato $\{X_i\}$, è già stata data nel § 2 (equazione 2.8). Riscriviamo queste equazioni introducendo la nuova idea che abbiamo intravisto nel § 3, cioè che l'evoluzione è condizionata da appropriati vincoli di non-equilibrio. Piú generalmente, vogliamo esprimere l'idea che l'evoluzione viene influenzata dalla variazione di qualche parametro presente nel problema, che può essere modificato dal mondo esterno. Chiamiamo queste quantità *parametri di controllo*, li indichiamo con λ, e scriviamo le equazioni 2.8 nella forma

$$\frac{\partial X_i}{\partial t} = F_i(\{X_j\}, \lambda) \qquad (2.12)$$

In un sistema dissipativo queste relazioni devono essere compatibili con le condizioni discusse alla fine del § 2. Le equazioni 2.9*b* forniscono un tipico esempio di queste *leggi di evoluzione* dissipative.

In considerazione della diversità dei fenomeni naturali, ci aspettiamo che la struttura delle funzioni $\{F_i\}$ dipenda in modo specifico dal si-

stema considerato e dal tipo di processo che si attua in questo sistema. Invero, non c'è a prima vista nessuna ragione perché il flusso del fluido contenuto fra i due piani nel problema di Bénard debba essere in alcun senso collegato alla aggregazione delle amebe *Dictyostelium discoideum* o alla glaciazione dell'èra Quaternaria! Tuttavia, nonostante questa sorprendente varietà, si possono selezionare alcune nozioni basilari che ci aiutano ad affrontare i fenomeni complessi in modo sistematico.

Una di queste è che, qualsiasi sia la forma delle F_i, in assenza di vincoli l'equazione 2.12 debba riprodurre lo stato di equilibrio. Poiché lo stato di equilibrio è uno stato stazionario, questo equivale a

$$F_i(\{X_{j,\,eq}\}, \lambda_{eq}) = 0 \qquad (2.13a)$$

Piú in generale, per uno stato stazionario di non-equilibrio,

$$F_i(\{X_{j,\,s}\}, \lambda) = 0 \qquad (2.13b)$$

Queste relazioni impongono alcune restrizioni. Per esempio, le leggi d'evoluzione devono essere tali che la positività della temperatura o le concentrazioni chimiche che vengono fuori come soluzioni, o la proprietà del bilancio dettagliato, siano identicamente soddisfatte. Questo è un punto importante, che mostra come l'analisi dei sistemi fisici non possa essere ridotta a un gioco matematico. I sistemi fisici si possono guardare per molti aspetti come altamente atipici, o non-generici, dal punto di vista matematico. Pensiamo per esempio ai sistemi conservativi hamiltoniani che sono costruiti con la condizione estremamente rigida che una particolare combinazione delle loro posizioni e velocità – l'energia – rimanga un invariante del moto.

Assicuratici che questa specificità dei sistemi fisici sia rispettata, cosa possiamo dire circa le possibili soluzioni delle equazioni 2.13*a* o 2.13*b*? È qui che il concetto di non-linearità comincia a giocare un ruolo predominante.

Facciamo un esempio. Sia X l'unica variabile di stato, λ il vincolo applicato, k un altro parametro. Possiamo facilmente immaginare un meccanismo tale che $A \rightleftarrows X \rightleftarrows D$ nella quale X evolve secondo la $dX/dt = \lambda - kX$ che porta a uno stato stazionario dato da

$$\lambda - kX_s = 0$$

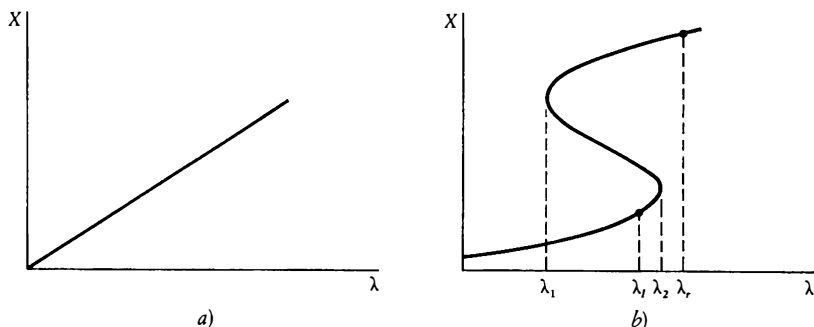
ovvero

$$X_s = \frac{\lambda}{k} \qquad (2.14)$$

La figura 2.5*a* rappresenta questa soluzione. Seguendo il modo in cui varia il valore dello stato stazionario X_s in termini del vincolo λ (che può essere, in particolare, la distanza dall'equilibrio), otteniamo una linea retta passante per l'origine. Questo è un tipico esempio di *legge lineare.* Per ogni valore λ del vincolo, vi corrisponde uno e uno solo stato stazionario X_s. Per di piú poiché due punti sono sufficienti per definire una linea retta, determinando sperimentalmente una coppia (λ_1, X_{1s}; λ_2, X_{2s}), possiamo prevedere in modo inequivocabile lo stato stazionario per ogni valore di λ. In altre parole, nel nostro sistema lineare il comportamento è obbligato a essere qualitativamente simile all'equilibrio, anche in presenza di forti vincoli di non-equilibrio.

Guardiamo ora la situazione illustrata nella figura 2.5*b*, che descrive una *legge non-lineare* che collega il valore dello stato stazionario X_s al parametro di controllo λ. La curva disegnata in questa figura funge naturalmente da esempio, poiché vi è un illimitato numero di possibili forme che descrivono dipendenze non-lineari. Quando λ è minore di λ_1, o piú grande di λ_2, la situazione è come quella della figura 2.5*a*. Ma per λ fra λ_1 e λ_2 il sistema può presentare parecchie soluzioni distinte. Questo cambiamento qualitativo avviene oltre una soglia critica, e permette già di concepire come una cella di Bénard possa essere destrogira o levogira, cosí come molti altri fatti sperimentali illustrati nel capitolo 1.

Una sorprendente differenza fra leggi lineari e non-lineari è la vali-

Figura 2.5
Effetto del parametro λ su una variabile di stato X.
a) Legge lineare.
b) Legge non-lineare.
Si noti la sensibilità della risposta X nelle vicinanze dei valori λ_1, λ_2 nel caso *b*).

dità o meno della proprietà della sovrapposizione. In un sistema lineare l'effetto di un'azione combinata di due diverse cause è solamente la sovrapposizione degli effetti di ogni causa presa separatamente. Ma in un sistema non-lineare una piccola causa può indurre effetti drammatici non proporzionati alla sua ampiezza. Ad esempio, supponiamo che nella figura 2.5*b* il parametro λ cambi prima da 0 a un valore subito alla sinistra del punto limite λ_2 della parte inferiore della curva (questa è la «causa prima»), e susseguentemente, un leggero aumento che lo porti subito alla destra di questo punto (questa è la «seconda causa»). Con $\lambda = 0$ come punto di partenza ognuna di queste cause che agisca da sola determina un «effetto» molto modesto, nella forma di un cambiamento molto lieve nel livello della variabile X. Ma sotto la loro azione combinata il sistema dovrà lasciare la parte inferiore della curva di risposta, poiché essa non è piú definita oltre λ_2. La sola via aperta è un salto discreto sulla parte superiore della curva. Questa è evidentemente una risposta molto forte, considerata la piccolezza della variazione di λ indotta dalla seconda causa.

È da notare che senza il mantenimento di una distanza appropriata dall'equilibrio, la non-linearità stessa non può dar luogo a soluzioni multiple. Infatti, all'equilibrio, il bilancio dettagliato introduce un'ulteriore condizione che restringe e persino fissa univocamente il valore della variabile di stato. Inoltre, per continuità questa univocità si estende ancora a un dominio di stati vicini all'equilibrio. È interessante illustrare questa proprietà con un semplice esempio che coinvolge le seguenti due reazioni accoppiate

$$A + 2X \underset{k_2}{\overset{k_1}{\rightleftarrows}} 3X \qquad (2.15)$$

$$X \underset{k_4}{\overset{k_3}{\rightleftarrows}} B$$

Qui la concentrazione x del prodotto X è l'unica variabile di stato, essendo chiaro che A e B vengono continuamente fornite dall'esterno o rimosse verso l'esterno per mantenere concentrazioni fisse, a e b. All'equilibrio, il bilancio dettagliato implica che

$k_1 a x^2 = k_2 x^3$ (l'equilibrarsi dei processi in avanti o a ritroso nella prima reazione)

$k_3 x = k_4 b$ (equilibrio della seconda reazione)

Queste reazioni fissano univocamente x_{eq} e impongono, inoltre, una condizione sulle concentrazioni a e b dei costituenti A e B:

$$x_{eq} = \frac{k_4 b_{eq}}{k_3} = \frac{k_1 a_{eq}}{k_2}$$

(2.16)

$$(b/a)_{eq} = \frac{k_1 k_3}{k_2 k_4}$$

D'altra parte, in uno stato stazionario lontano dall'equilibrio, è sufficiente cancellare l'effetto globale delle due reazioni dirette con quello delle reazioni a ritroso, avendo cosí

(2.17) $$-k_2 x_s^3 + k_1 a x_s^2 - k_3 x_s + k_4 b = 0$$

Questa è un'equazione cubica in x_s e può avere fino a tre soluzioni per certi valori di a e b, come nel caso della figura 2.5b e diversamente a ciò che succede all'equilibrio. Possiamo pertanto dire che il non-equilibrio rivela le potenzialità nascoste nelle non-linearità, che rimangono «assopite» vicino o all'equilibrio.

La possibilità di avere soluzioni multiple in un sistema non-lineare, solleva automaticamente la questione della *scelta* fra i diversi risultati: chiaramente, una data cella di Bénard non può essere sia destrogira che levogira. Per capire questo aspetto cruciale sono necessari ulteriori concetti, che verranno esaminati nei paragrafi seguenti. Prima di chiudere questo paragrafo, vogliamo dare una lista di alcuni importanti esempi di non-linearità che si incontrano ripetutamente in vari campi.

Meccanica dei fluidi. La meccanica dei fluidi presenta numerosi problemi di non-linearità, essenzialmente perché la velocità di cambiamento di una data proprietà dipende, fra altri fattori, dal trasporto di questa proprietà dal naturale flusso del fluido, la cui velocità $\boldsymbol{v}$ è una delle variabili del problema. Cosí, l'equazione della conservazione del momento coinvolge il termine evidentemente non-lineare $\boldsymbol{v} \cdot \nabla \boldsymbol{v}$. Inoltre, in molti problemi anche la temperatura T e la densità di massa ρ sono fra le variabili sconosciute. Non-linearità addizionali sorgono allora attraverso l'accoppiamento di queste variabili alla velocità di convezione, dando luogo ai termini $\nabla (\rho \boldsymbol{v})$ e $\boldsymbol{v} \cdot \nabla T$ rispettivamente nelle equazioni di equilibrio della massa e dell'energia interna.

Chimica e biologia. In chimica e biologia le non-linearità possono nascere intrinsecamente, indipendentemente dalle disomogeneità spaziali. La maggior parte delle reazioni chimiche sono il risultato di collisioni. Per esempio, $X + Y \rightarrow D$ implica che le molecole delle specie X e Y devono interagire, e questo dà luogo a una velocità proporzionale a $c_X c_Y$, cioè a una funzione non-lineare. Sono di particolare interesse, tuttavia, come già sottolineato nei §§ 4 e 7, le non-linearità collegate alla *regolazione*. Qui una sostanza X può attivare o inibire la velocità della propria produzione o della produzione di un altro costituente, che a sua volta si nutre della prima sostanza. Nella chimica ordinaria due caratteristici esempi sono l'autocatalisi, illustrata dal primo passo della 2.15, e il *feedback* termico discusso nel capitolo 1, § 4 (cfr. equazione 1.5 e figura 1.11). In biologia la regolazione è intimamente legata alla peculiare struttura e reattività degli enzimi. In molti casi dà luogo a non-linearità non-polinomiali della forma:

$$\left(\frac{dc}{dt}\right)_{reg} \sim \frac{ac^m}{1 + bc^n}$$

che dà attivazione se $m \geqslant n$ e inibizione se $m < n$. Si incontrano simili espressioni matematiche in altri campi, come nei circuiti elettrici che contengono elementi attivi o nei processi radiativi nella dinamica climatica.

5. *I molti aspetti della seconda legge.*

Abbiamo visto che la non-linearità combinata con i vincoli del non-equilibrio dà origine a soluzioni multiple e dunque alla diversificazione dei comportamenti presentati da un sistema. Fino a questo momento abbiamo guardato a queste soluzioni una volta stabilitesi, per una qualche ragione, e ne abbiamo analizzato qualche caratteristica. È ora tempo di sollevare la questione di *come* queste soluzioni delle leggi di evoluzione, equazione 2.12, si ottengono nel corso del tempo.

Per i sistemi conservativi la risposta a questa questione è notevolmente semplice, almeno in principio. L'evoluzione è interamente dettata dalle condizioni iniziali (cfr. le figure 1.8 e 2.2), che fissano le varie costanti del moto e dunque le traiettorie stesse, attraverso relazioni come l'equazione 2.6. In pratica, tuttavia, questa ricetta funziona in mo-

do utile solo per situazioni relativamente semplici. Posticipiamo la discussione di altri casi al capitolo III.

È nel caso dei sistemi dissipativi che la ricerca di criteri per l'evoluzione porta a qualche risultato profondo e inaspettato. Infatti, la non-invarianza delle equazioni per le variabili di stato rispetto all'inversione del tempo, suggerisce che tali sistemi possano essere caratterizzati da un approccio *irreversibile* verso uno stato finale, che, in generale, ci si aspetta essere un regime indipendente dal tempo.

Per sistemi isolati, nei quali non è permesso alcuno scambio con l'ambiente, questa tendenza irreversibile viene espressa con la celebre seconda legge della termodinamica. Secondo questa legge esiste una funzione delle variabili di stato del sistema (generalmente detta *funzione di stato*), monotòna durante l'avvicinarsi all'unico stato finale di equilibrio termodinamico. Generalmente per questa funzione viene scelta l'entropia S, una quantità introdotta prima fenomenologicamente nello studio delle macchine termiche e studiata in seguito a un livello piú microscopico da Boltzmann, Gibbs e altri. La formulazione della seconda legge è quindi

$$\frac{dS}{dt} \geqslant 0 \qquad \text{(sistema isolato)} \tag{2.18}$$

Questo viene generalmente interpretato come una tendenza al disordine, poiché la teoria microscopica associa l'entropia al numero di stati che si possono realizzare, date le condizioni applicate al sistema. Si può allora sostenere che piú è ristretto il numero di stati, piú ordinato sarà il sistema. Per esempio, un solido vicino allo zero assoluto ha una bassa entropia, mentre un gas a temperatura e pressione ordinarie in cui le particelle si muovono veloci e collidono in tutte le possibili direzioni, possiede un'alta entropia.

La suddetta interpretazione dell'entropia e della seconda legge è piuttosto ovvia per sistemi costituiti da particelle debolmente interagenti, ai quali erano originariamente rivolte le argomentazioni di Boltzmann. Per esempio, sia gli esperimenti che la normale intuizione suggeriscono che una preparazione iniziale che consiste nel riempire solo mezza scatola di materia (figura 2.6*a*) evolverà verso una situazione nella quale la materia è distribuita piú o meno uniformemente attraverso la scatola (vedi figura 2.6*b*). In una tale evoluzione il disordine è chiaramente in aumento. In concomitanza, poiché il volume nel caso (*a*) è metà del volume occupato nel caso (*b*), il numero degli stati permessi è anche aumentato. Cosí, se definiamo l'entropia come la misura del nu-

mero di stati, realizziamo che l'evoluzione illustrata nella figura 2.6 è riassunta dalla relazione 2.18, che indica un'irreversibile tendenza al massimo disordine.

Per sistemi fortemente interagenti la suddetta interpretazione non regge in maniera diretta, poiché sappiamo che per questi sistemi c'è possibilità di evolvere verso stati piú ordinati attraverso un meccanismo di transizioni di fase.

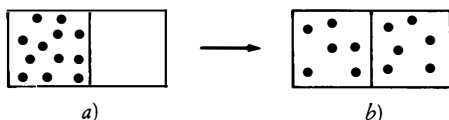
Per quanto dettagliata sia la definizione microscopica di entropia per questi sistemi, ci aspettiamo che in assenza di vincoli esterni lo stato del sistema – l'equilibrio termodinamico – sia totalmente dominato dalle interazioni intermolecolari. Ne segue che in un sistema fisico isolato in cui le interazioni gravitazionali non sono la causa principale della dinamica, dopo un transiente iniziale descritto dalla relazione 2.18 durante la quale il sistema è dominato da processi irreversibili, si raggiunge uno stato finale di equilibrio, nel quale la scala caratteristica di ordine spaziale sarà necessariamente microscopica, dell'ordine di pochi Å. Un esempio caratteristico di questo tipo di ordine è il cristallo, in cui le dimensioni della zona reticolare sono in effetti dell'ordine del diametro molecolare. Chiaramente una cella di Bénard o un orologio chimico non si formeranno mai spontaneamente in un tale stato.

Guardiamo ora i sistemi non isolati, che scambiano energia o materia con l'ambiente. La variazione di entropia sarà ora la somma di due termini, uno dovuto a questi scambi, indicato con d_eS (flusso di entropia) e uno dovuto ai fenomeni che si svolgono nel sistema, indicato con d_iS (produzione di entropia):

$$\frac{dS}{dt} = \frac{d_iS}{dt} + \frac{d_eS}{dt} \tag{2.19a}$$

Per un sistema isolato $d_eS = 0$, e l'equazione 2.19*a* assieme alla 2.18 si riduce a $dS = d_iS \geqslant 0$, che è l'usuale formulazione della seconda legge. Se il sistema non è isolato, d_iS descriverà quei processi (irreversibili) che

Figura 2.6

Spontanea omogeneizzazione che ha luogo quando un gas inizialmente contenuto nella parte sinistra di una scatola *a*) viene lasciato libero di occupare l'intero volume *b*). Durante questa evoluzione crescono sia l'entropia che il disordine.

si verificherebbero anche in assenza del termine di flusso d_eS. Necessitiamo dunque della seguente formulazione piú estesa della seconda legge:

$$\frac{d_iS}{dt} \geqslant 0 \qquad \text{(sistema non isolato)} \tag{2.19b}$$

Finché d_iS è strettamente positivo, i processi irreversibili continueranno a verificarsi entro il sistema. Cosí, $d_iS > 0$ è equivalente alla condizione di dissipazione discussa nel § 2. Se, d'altra parte, d_iS è zero, il processo sarà reversibile e dovrebbe unicamente collegare i vicini stati di equilibrio attraverso una lenta variazione del termine di flusso d_eS. È quest'ultima situazione che viene abitualmente studiata nei corsi introduttivi di termodinamica, dedicata ai cicli di Carnot o ad altri tipi di macchine termiche.

Fra i piú comuni processi irreversibili che contribuiscono a d_iS vi sono le reazioni chimiche, la conduzione del calore, la diffusione, la dissipazione viscosa, o i fenomeni di rilassamento nei sistemi polarizzati elettricamente o magneticamente. Per ciascuno di questi fenomeni si definiscono un appropriato flusso interno, J_k, che denota essenzialmente la sua rapidità, e una forza guida X_k, collegata al mantenimento del vincolo di non-equilibrio. Un aspetto piú interessante è che d_iS diventa una forma quadratica di J_k e X_k, almeno per i fenomeni che si possono ricondurre a una descrizione locale simile a quella data dall'idrodinamica o dalla cinetica chimica:

$$\frac{d_iS}{dt} = \sum_k J_k X_k \tag{2.20}$$

La tabella 2.1 riassume i flussi e le forze associate a qualche fenomeno irreversibile comunemente osservato.

In generale, i flussi J_k sono funzioni molto complicate delle forze. Si ottiene una situazione particolarmente semplice quando la relazione è lineare

$$J_k = \sum_l L_{kl} X_l \tag{2.21}$$

in cui L_{kl} indica l'insieme dei *coefficienti fenomenologici*.
Questo è ciò che avviene vicino all'equilibrio dove, in aggiunta, $L_{kl} = L_{lk}$, come mostrato da Lars Onsager sulla base di argomentazioni collegate alla validità del bilancio dettagliato all'equilibrio. È da notare, tuttavia,

che certi stati lontani dall'equilibrio si possono ancora caratterizzare da una dipendenza lineare della forma 2.21 sia per cause accidentali sia per la presenza di speciali tipi di processi regolatori. In ogni caso, sostituendo ai flussi che appaiono nelle equazioni di evoluzione (cfr. equazione 2.12) le loro funzioni in termini delle forze, e notando che queste ultime possono essere a loro volta espresse in termini delle stesse variabili di stato, arriviamo in un modo alternativo alla conclusione già formulata nel § 4, e cioè che vicino all'equilibrio le eventuali non-linearità presenti in un sistema diventano inefficaci.

Torniamo ora alla forma generale (equazione 2.19*a*). Abbiamo già sottolineato che la seconda legge impone $d_iS \geqslant 0$. In contrapposizione a questa restrizione, tuttavia, in un sistema non-isolato ($d_eS \neq 0$), non esiste legge fisica che imponga il segno di d_eS. In altre parole, d_eS può essere positivo o negativo, a seconda del sistema considerato. È dunque concepibile che d_eS diventi sufficientemente negativo ed ecceda in valore assoluto la grandezza di d_iS, nel qual caso certi stati di evoluzione sono caratterizzati da

$$dS/dt < 0 \qquad \text{(sistema non isolato)} \tag{2.22}$$

Secondo la tradizionale interpretazione di entropia, questo significa che, grazie al flusso di entropia, il disordine è diminuito nel corso dell'evoluzione. Questo aprirebbe la via alla comprensione dell'origine dei fenomeni complessi visti nel capitolo III su base termodinamica.

La situazione non è tuttavia semplice. Come già detto nella prima parte di questo paragrafo, la connessione fra entropia e ordine è chiara

Tabella 2.1.
Flussi e forze irreversibili comunemente osservati*.

Fenomeno	Flusso	Forza	Tipo di vettore
Conduzione termica	Flusso di calore, J_{th}	grad $(1/T)$	Vettore
Diffusione	Flusso massiccio di costituenti, i, $J_{d,i}$	$-[\text{grad}\,(\mu_i/T) - F_i]$	Vettore
Flusso viscoso	Parte dissipativa del tensore di pressione, P	$(1/T)$ grad v	Tensore (2° grado)
Reazione chimica	Velocità di reazione, ρ, w_ρ	Affinità della reazione divisa per T: $\mathcal{A}_\rho/T$	Scalare

* T = temperatura; μ_i = potenziale chimico del costituente i; F_i = forza esterna per unità di massa agente su i; v velocità idrodinamica. L'affinità $\mathcal{A}_\rho$ è legata a μ_i secondo la relazione $\mathcal{A}_\rho = -\sum_i \nu_{i\rho}\mu_i$ in cui il coefficiente stechiometrico $\nu_{i\rho}$ fornisce il numero di molecole prodotte ($\nu > 0$), o distrutte ($\nu < 0$), nella reazione.

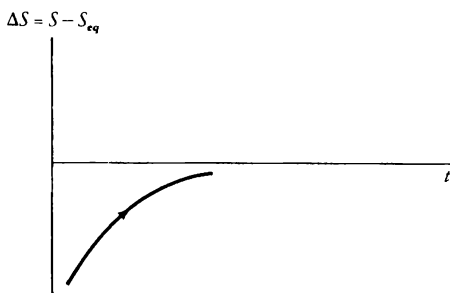
nei sistemi ideali (non-interagenti), o nei sistemi piú generali all'equilibrio, in cui la scala spaziale caratteristica delle strutture ordinate che si formano è paragonabile alla scala delle forze intermolecolari. D'altra parte, l'ordine associato con l'apparire di strutture dissipative sotto condizioni di non-equilibrio non si può collegare in modo semplice a una diminuzione di entropia. Dovremmo cosí considerare l'affermazione della 2.22 come niente piú che una forte motivazione per la ricerca di una piú sottile e inequivocabile caratterizzazione della complessità.

6. *Stabilità*.

Descriviamo ora da un altro punto di vista il contenuto della seconda legge nel caso di un sistema isolato. Da una parte, si asserisce che, qualsiasi sia la preparazione iniziale di un sistema, si raggiungerà nel tempo uno stato unico – lo stato di equilibrio. Dall'altra parte, questa tendenza implicherà un aumento monotòno di entropia nel tempo, o, alternativamente, un aumento monotòno dell'eccedenza $\Delta S = S - S_{eq}$ da valori negativi a zero, come mostrato nella figura 2.7.

Entrambe le asserzioni sono in effetti equivalenti, e in linguaggio piú matematico, esse significano che l'equilibrio è un sistema *asintoticamente e globalmente stabile* e che l'entropia è una *funzione di Ljapunov* che esprime questa forma estremamente forte di stabilità.

Figura 2.7

Un punto di vista alternativo della seconda legge della termodinamica per un sistema isolato l'eccesso di entropia ΔS è negativo e cresce monotonamente nel tempo, giocando cosí il ruolo di una funzione di Ljapunov. Lo stato di equilibrio è asintoticamente stabile e costituisce, infatti, un attrattore globale.

Fermiamoci un attimo per definire piú precisamente questi termini, che giocheranno un ruolo importante nella nostra successiva analisi (vedere anche il capitolo 1, §§ 3 e 4, nei quali la stabilità asintotica era già stata invocata in modo intuitivo).

Consideriamo un sistema – dissipativo o conservativo – che evolve secondo certe leggi dinamiche, date, per esempio, dall'equazione 2.12. A causa di un meccanismo che qui non occorre specificare, questo sistema si è trovato in un particolare stato X_S, nel quale poi è rimasto. In un sistema conservativo X_S può essere uno stato di equilibrio, mentre in un sistema dissipativo può rappresentare lo stato di equilibrio termodinamico o uno stato stazionario di non-equilibrio.

Nella realtà, un sistema non rimane mai in un preciso stato al variare del tempo. Per cominciare, la maggior parte dei sistemi sono in contatto con un ambiente complesso o anche imprevedibile. Questo ambiente continua a comunicare loro piccole (o raramente grandi) quantità di materia, momento ed energia, con il risultato che è praticamente impossibile controllare alcuna variabile di stato con una precisione illimitata. Ogni scienziato sperimentale ha provato questa frustrazione in minore o maggior misura, e i termini frequentemente usati «errore sperimentale» o «intervallo di tolleranza» riflettono unicamente l'interferenza dell'ambiente con la dinamica intrinseca. Esprimeremo questo onnipresente fatto dicendo che lo stato istantaneo del nostro sistema non è X_S ma piuttosto uno stato vicino X collegato a X_S attraverso la

$$X(t) = X_S + x(t) \tag{2.23}$$

La quantità x sarà chiamata *perturbazione*.

C'è un secondo aspetto piú importante di questo stesso problema. Molti sistemi con cui abbiamo a che fare sono composti di un grande numero di entità interagenti. Per questi sistemi le variabili di stato di cui abbiamo parlato in questo capitolo devono essere viste in senso statistico: esse rappresentano o le medie degli stati istantanei su un lungo intervallo di tempo, o forse il piú probabile fra i valori che possono assumere le variabili di stato istantanee. Nel mondo reale dunque, una registrazione dello stato istantaneo mostrerà continue deviazioni dai valori delle variabili di stato macroscopiche che vengono generati spontaneamente dal sistema indipendentemente dall'ambiente. Chiamiamo queste intrinseche deviazioni *fluttuazioni*. Il loro effetto si può ancora esprimere con l'equazione 2.23, sempre che sia chiaro che $x(t)$ fa ora parte della dinamica del sistema considerato.

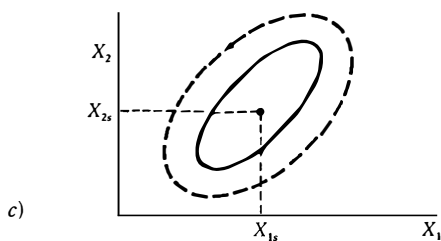
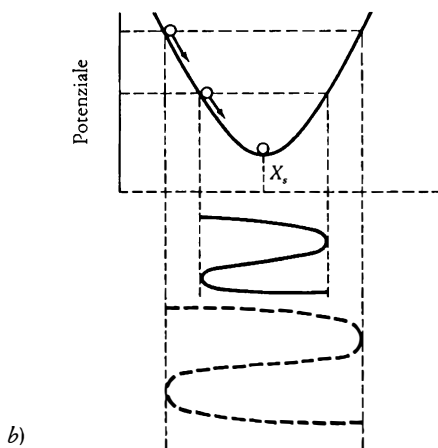
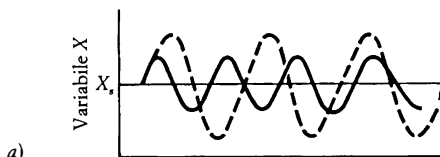
Figura 2.8

Tre aspetti della stabilità di Ljapunov.

a) Le perturbazioni attorno allo stato di riferimento X_S rimangono limitate per ogni valore del tempo.

b) Visualizzazione meccanica di *a*), nella quale il moto di un punto materiale in un pozzo di potenziale viene rappresentato per due diverse condizioni iniziali.

c) Le orbite dello spazio delle fasi corrispondenti a due diverse perturbazioni dallo stato di riferimento (X_{1S}, X_{2S}).

Quest'ultima figura illustra anche la nozione di stabilità orbitale.

Tipicamente, si ha una fluttuazione sotto forma di evento localizzato e di piccola scala. Per esempio, in una piccola parte del sistema macroscopico, certe particelle possiedono una grande velocità termica in confronto alle loro vicine. Poco dopo, la densità delle particelle in un altro piccolo subvolume può essere inferiore al valore delle zone limitrofe, e cosí via. Possiamo facilmente capire che tali eventi saranno essenzialmente casuali. Termini come «rumore di fondo», spesso usati per indicare le fluttuazioni, illustrano questa proprietà.

Abbiamo spiegato le ragioni per cui lo stato di un sistema devierà continuamente dallo *stato di riferimento* X_S. Discutiamo ora la risposta del sistema a queste deviazioni. I casi seguenti costituiscono una lista completa di tutte le possibilità.

Caso 1. Lo stato $X(t)$ rimane nelle vicinanze dello stato di riferimento X_S per ogni istante piú grande di un tempo iniziale t_0. Piú precisamente, per ogni prescritto $\varepsilon > 0$, possiamo trovare una perturbazione iniziale piú piccola di δ, dove δ dipende in generale da ε e da t_0, cosicché la deviazione $x(t)$ rimane piú piccola di ε per ogni istante maggiore di t_0. Diciamo allora che lo stato X_S è *stabile nel senso di Ljapunov*. La stabilità di Ljapunov è già stata illustrata nella parte sinistra, riquadro inferiore, della figura 1.8. La figura 2.8 fornisce tre punti di vista alternativi di questa nozione.

In molti casi non interessa la risposta di un certo stato alle perturbazioni, ma piuttosto la risposta di una sequenza intera di stati che definiscono una traiettoria. Si introduce allora una misura appropriata della distanza fra riferimento e traiettorie perturbate e si applica a essa la suddetta definizione di stabilità. Questo porta al concetto di *stabilità orbitale*.

Caso 2. Lo stato $X(t)$ torna a X_S quando il tempo tende a ∞; in altre parole, la perturbazione $x(t)$ va a zero nel tempo. Diciamo allora che X_S è *asintoticamente stabile*. La stabilità asintotica viene illustrata a destra, nel riquadro inferiore della figura 1.8, e ulteriori aspetti sono riassunti nella figura 2.9. La definizione si può estendere alla *stabilità asintotica orbitale*. Possiamo facilmente capire come la stabilità asintotica implichi necessariamente irreversibilità e per questa ragione non si può applicare ai sistemi conservativi. Questi ultimi possono al piú essere stabili nel senso di Ljapunov.

D'altra parte, i sistemi dissipativi sono in grado di eliminare l'effetto delle perturbazioni che possono agire su di essi e ristabilire in questo modo lo stato di riferimento. Questo assicura la prevedibilità e la riproducibilità di questo regime, a cui ci si riferirà d'ora in avanti col termine *attrattore*. Per esempio, le celle di Bénard, le oscillazioni nei reagenti della BZ, o lo stato di equilibrio termodinamico in un sistema isolato sono attrattori di sistemi dinamici appropriatamente definiti. Senza dubbio, la stabilità asintotica è una delle manifestazioni piú sorprendenti del ruolo costruttivo dell'irreversibilità in natura.

Caso 3. Lo stato $X(t)$ non rimane nelle vicinanze di X_S, nel senso che non esiste una perturbazione iniziale per cui la grandezza della perturbazione $x(t)$ rimane inferiore a un prescritto valore e per ogni istante maggiore di t_0. Diciamo allora che lo stato di riferimento X_S è *instabile*. Questa situazione si manifesterà con uno stadio iniziale di rapida (spesso esponenziale) crescita della perturbazione. Tutti questi aspetti si estendono facilmente per definire l'*instabilità orbitale*.

Una manifestazione di instabilità si avrebbe attraverso un diagramma simile a quello della figura 2.9*b*, ma con le frecce invertite, indicando il movimento nella direzione opposta. Un'altra forma di instabilità è illustrata nella figura 2.10.

Gli stati di riferimento instabili si possono avere sia nei sistemi conservativi che dissipativi. Due esempi sono forniti dallo stato di equili-

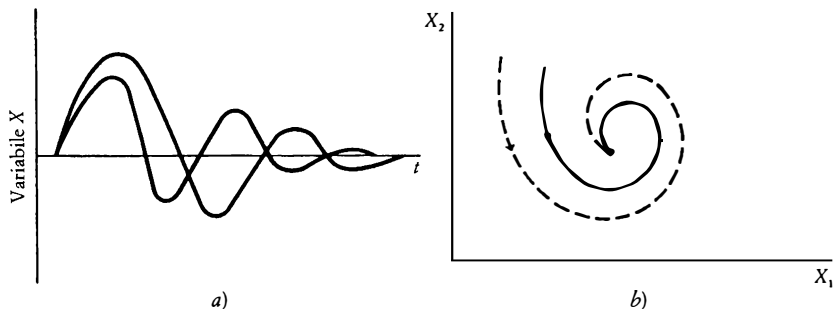
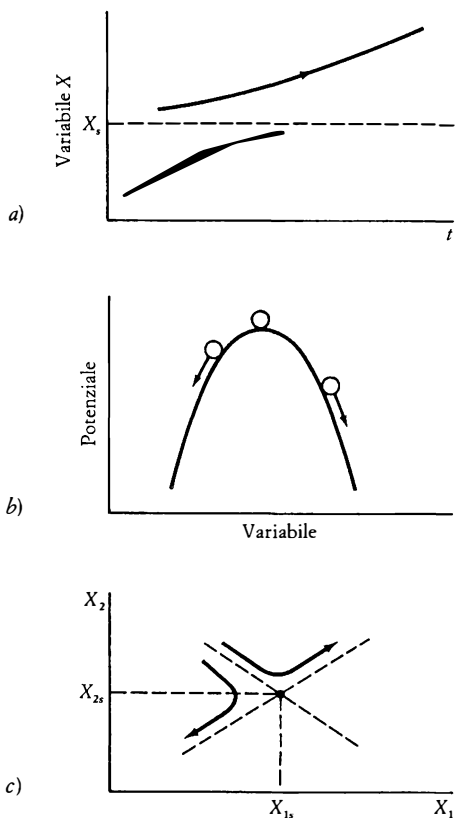
Figura 2.9
Due esempi di stabilità asintotica.
a) Evoluzione temporale della variabile di stato X.
b) Traiettorie dello spazio delle fasi che convergono verso un attrattore.

brio sulla verticale superiore nel pendolo e dallo stato di riposo nel problema di Bénard oltre la soglia di convezione. Tuttavia, mentre non si applica nessuna restrizione al caso dissipativo, in un sistema conservativo lo stato rimane instabile per inversione temporale. Ciò è compatibile con la figura 2.10, ma in contraddizione con la figura 2.9*b* nella quale le frecce fossero invertite.

Figura 2.10

Tre esempi dell'instabilità di uno stato di riferimento.

a) Evoluzione temporale di una variabile di stato X.

b) Un esempio meccanico fornito dal moto di un punto materiale inizialmente posto in cima a un rilievo.

c) Rappresentazione dello spazio delle fasi.

Caso 4. Lo stato $X(t)$ rimane nelle vicinanze dello stato di riferimento X_S per valori delle perturbazioni iniziali inferiori a una certa soglia, ma se ne allontana per perturbazioni oltre questa soglia. Diciamo allora che X_S è *localmente stabile*, ma *globalmente instabile*. Se, d'altra parte, la stabilità prevale per ogni valore della perturbazione iniziale, parliamo di *stabilità globale*, essendo lo stato X_S, in questo caso, un *attrattore globale*. Per esempio, l'equilibrio termodinamico in un sistema isolato è un attrattore globale, ma uno stato di convezione nel problema di Bénard, nel quale una particolare cella (diciamo la 1 nella figura 1.3) è destrogira, è globalmente instabile rispetto a una perturbazione finita che produce un flusso levogiro nella stessa cella. Una rappresentazione meccanica di stabilità locale e di instabilità globale è data dalla figura 2.11. Un punto mobile con una piccola quantità di energia cinetica che inizia a partire (senza attrito) vicino alla valle (1) rimarrà entro di essa, ma lo stesso oggetto inizialmente dotato di una grande quantità di energia cinetica supererà il rilievo (2) e seguirà successivamente un tragitto che comprende entrambe le valli (1) e (2). In altre parole, le posizioni di equilibrio lungo la valle (1) (e similmente per la valle 2)) sono stabili solo nei riguardi di perturbazioni inferiori all'altezza del rilievo.

Alla luce della nostra discussione sulla stabilità, possiamo ora comprendere meglio il passo concettualmente piú avanzato ottenuto dal passaggio dai sistemi isolati, descritti dalla relazione 2.18, ai sistemi aperti lontani dall'equilibrio, descritti dalle relazioni 2.19 e 2.20. In un sistema isolato l'equilibrio è globalmente stabile, e questo viene espresso dall'esistenza di una funzione, l'entropia o, piú precisamente, l'eccedenza di entropia $\Delta S = S - S_{eq}$, che gode delle proprietà (cfr. figura 2.7):

$$(2.24) \qquad \Delta S \leqslant 0, \qquad \frac{d}{dt}\Delta S \geqslant 0 \qquad \text{(sistema isolato)}$$

Chiamiamo ΔS una *funzione di Ljapunov*. Nei testi di matematica si dimostra che la 2.24 è effettivamente equivalente alla stabilità globale.

In un sistema non isolato, non all'equilibrio, a causa dell'esistenza dei due termini d_eS e d_iS nel bilancio dell'entropia, la seconda legge $d_iS/dt \geqslant 0$, non impone piú il segno della variazione dell'entropia (né, infatti, di ogni altra funzione di stato). Cosí non c'è una funzione di Ljapunov universale, e questo *solleva il problema della stabilità* degli stati lontani dall'equilibrio. La possibilità della perdita della stabilità sotto certe condizioni apre la via a fenomeni di transizione che possono portare a differenziazioni interne e a un comportamento complesso.

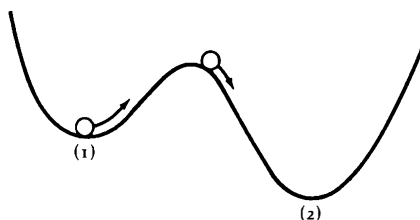
Vale tuttavia la pena di notare un'eccezione, ovvero i sistemi aperti vicino all'equilibrio. A causa della linearità delle leggi fenomenologiche e della simmetria dei coefficienti in questo ambito (cfr. equazione 2.21), la produzione di entropia d_iS/dt (equazione 2.20), diventa una forma quadratica definita positiva dei vincoli X_k. È per di piú possibile mostrare che ogni deviazione dallo stato stazionario porta a una evoluzione temporale tendente a diminuire il valore di d_iS/dt. In altre parole, abbiamo una situazione simile all'equazione 2.24, nella quale la produzione di entropia gioca il ruolo di una funzione di Ljapunov che garantisce la stabilità globale degli stati stazionari vicini all'equilibrio.

Si deve sottolineare che, nonostante l'assenza di una legge generale che garantisce la stabilità degli stati lontani dall'equilibrio, è ancora possibile trovare criteri utili che portano a proprietà piú deboli, come, per esempio, le *condizioni sufficienti* per la stabilità. Uno di tali criteri è fornito dalla condizione di positività dell'eccesso di produzione di entropia valutata attorno a uno stato di riferimento di non-equilibrio.

7. *Biforcazione e rottura della simmetria.*

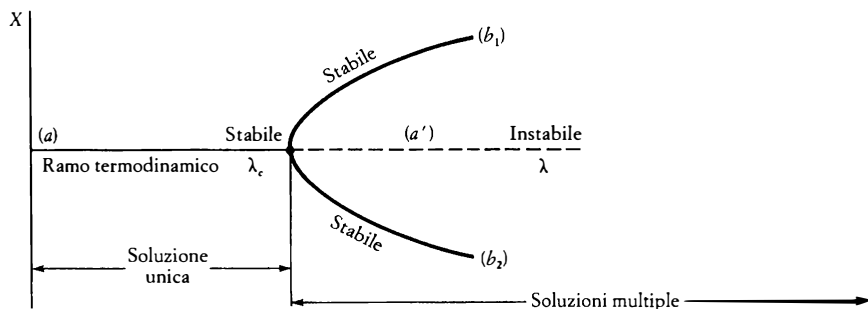
Consideriamo ancora una volta il problema di Bénard da un diverso punto di vista. Abbiamo già insistito sull'importanza del vincolo che porta il sistema lontano dall'equilibrio e sul ruolo delle non-linearità. Sappiamo anche che le diverse transizioni osservate sperimentalmente hanno luogo quando questo parametro viene modificato dallo sperimentatore. Vogliamo estrapolare alcuni aspetti generali che sono presenti in molti altri fenomeni di transizione nei sistemi fisico-chimici. A

Figura 2.11

Stabilità globale contro stabilità locale. Un punto mobile che parte dalla valle (1) può rimanere entro di essa o giungere alla valle (2) a seconda che la sua energia cinetica iniziale sia minore o maggiore di una soglia finita.

questo fine rappresentiamo su un grafico (cfr. figura 2.12) il modo in cui la componente verticale della velocità del flusso di un certo punto viene influenzato dal vincolo o, in un linguaggio piú generale, il modo in cui la variabile di stato del sistema X (o $x = X - X_S$) è influenzata dal parametro di controllo λ. Otteniamo cosí un grafico noto come *diagramma di biforcazione.*

Per piccoli valori di λ è possibile solo una soluzione: lo stato di riposo nell'esperimento di Bénard. È la diretta estrapolazione dell'equilibrio e divide con essa l'importante proprietà della stabilità asintotica, poiché in questa situazione il sistema è in grado di smorzare le fluttuazioni interne o i disturbi esterni. Per questo motivo, chiamiamo questo ramo degli stati il *ramo termodinamico.* Ma oltre un valore critico, indicato con λ_c nella figura 2.12, troviamo che gli stati su questo ramo diventano instabili: l'effetto delle fluttuazioni o delle piccole perturbazioni esterne non viene smorzato ulteriormente. Il sistema agisce come un amplificatore, si allontana dallo stato di riferimento ed evolve verso un nuovo regime, lo stato di convezione nell'esperimento di Bénard. I due regimi si fondono per $\lambda = \lambda_c$, ma si differenziano per $\lambda > \lambda_c$. Questo è il fenomeno della *biforcazione.* Possiamo facilmente comprendere come mai questo fenomeno debba essere associato ai cambiamenti catastrofici e ai conflitti. Invero, al momento cruciale della transizione (nelle vicinanze di $\lambda = \lambda_c$), il sistema deve compiere una scelta critica che, nel problema di Bénard, viene associata alla comparsa di una cella de-

Figura 2.12

Diagramma di biforcazione che mostra come una variabile di stato X sia influenzata dalla variazione del parametro di controllo. L'unica soluzione *a*), il ramo termodinamico, perde la sua stabilità per $\lambda = \lambda_c$. Per questo valore del parametro di controllo si generano due nuovi rami di soluzione (b_1, b_2) che sono stabili nell'esempio scelto.

strogira o levogira in una certa regione dello spazio (figura 2.12, rami b_1 o b_2). Nella descrizione dell'apparato sperimentale nulla permette all'osservatore di stabilire in anticipo lo stato che verrà scelto; deciderà solamente il caso, attraverso la dinamica delle fluttuazioni.

Il sistema ispezionerà il terreno, farà alcuni tentativi, forse infruttuosi all'inizio, e finalmente adotterà una particolare fluttuazione. Con il realizzarsi di questa fluttuazione, il sistema diventa un oggetto storico, nel senso che la sua evoluzione susseguente dipenderà da questa scelta critica.

In questa descrizione siamo dunque riusciti a formulare, in termini astratti, l'importante interazione fra caso e vincoli, fra fluttuazioni e irreversibilità, che era alla base della maggior parte dei fenomeni esaminati nel capitolo 1. È da notare la similitudine fra queste idee e la nozione di mutazione e di selezione familiare della evoluzione biologica. Si può infatti dire che le fluttuazioni sono la controparte fisica della mutazione, mentre la stabilità gioca il ruolo della selezione. Perfino la stessa struttura di un diagramma di biforcazione ricorda gli alberi filogenetici abbondantemente usati in biologia.

Se anziché al flusso di Bénard fossimo interessati alla reazione BZ, alla aggregazione del *Dictyostelium discoideum*, o praticamente a qualsiasi altro fenomeno di transizione, e se costruissimo per questi sistemi un diagramma di biforcazione, troveremmo la stessa struttura qualitativa. Le uniche differenze riguarderebbero la terminologia, che è necessariamente specifica per ogni problema trattato. Troviamo cosí una profonda unità che ci permette di asserire che il passaggio verso la complessità è intimamente collegato alla *biforcazione* di nuovi rami di soluzioni che nascono dall'*instabilità* di uno stato di riferimento causata dalle non-linearità e dai vincoli che agiscono su un sistema aperto.

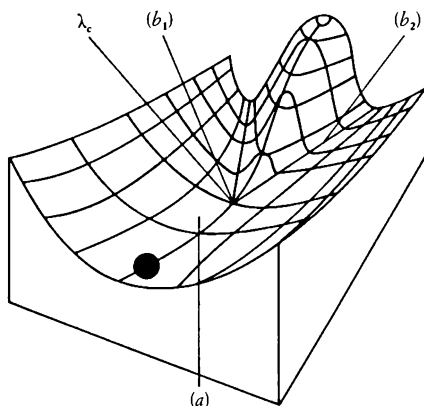
La figura 2.13 rappresenta un analogo meccanico del fenomeno. Una palla si muove in una valle, equivalente al ramo *a* della figura 2.12, che a un certo punto λ_c si ramifica e porta a due nuove valli, i rami b_1 e b_2, separati da una collina. Benché sia troppo presto per fare analogie ed estrapolazioni a questo punto, è stimolante immaginare per un momento che invece della palla, nella figura 2.13, vi sia un dinosauro prima della fine dell'èra Secondaria, o un gruppo di nostri antenati che stanno per decidere tra la scrittura ideografica e quella simbolica.

Abbiamo ripetutamente sottolineato il fatto che la biforcazione è una fonte di innovazione e diversificazione, poiché dota un sistema di nuove soluzioni. Ma quali sono le caratteristiche di queste soluzioni? Posticipiamo la parte principale della risposta a questa domanda al capi-

tolo III. Chiudiamo questo paragrafo rimarcando un punto particolare già intravisto nella discussione dei fatti sperimentali del capitolo I, e cioè che la biforcazione genera soluzioni che mostrano *rotture di simmetria*.

È un dato sperimentale che certi fenomeni fisici si possono ricondurre esattamente alla stessa descrizione anche quando gli esperimenti vengono eseguiti sotto diverse condizioni. Come abbiamo visto nel capitolo I, §§ 3 e 9, per esempio, un liquido prima della soglia per la convezione di Bénard ha identiche proprietà per tutti gli osservatori disposti lungo una certa direzione (invarianza traslazionale). Un altro esempio familiare preso dalla meccanica è la caduta libera di un corpo vicino alla superficie della Terra; essa avviene lungo la verticale, non solo per un osservatore solidale con la Terra, ma per tutti gli osservatori che si muovono rispetto a questi con velocità costante (invarianza galileiana).

Questi sono fatti importanti perché impongono dei vincoli che devono essere rispettati da ogni ragionevole modello matematico. In particolare, le leggi che descrivono i fenomeni devono mostrare *simmetrie* che garantiscano l'*invarianza* dei fenomeni stessi. Queste simmetrie devono essere soddisfatte per tutte le operazioni nello spazio e nel tempo che trasformano in se stesse gli osservatori per i quali è vera l'invarianza. Molto frequentemente queste operazioni costituiscono un oggetto matematico noto come gruppo, poiché le operazioni eseguite consecuti-

Figura 2.13
Esempio meccanico del fenomeno della biforcazione.

vamente sono equivalenti a una operazione che è essa stessa compatibile con l'invarianza. Cosí, le equazioni di fluidodinamica che descrivono un liquido riscaldato dal basso e contenuto fra due piani infinitamente estesi nella direzione orizzontale sono invarianti rispetto al gruppo delle traslazioni lungo un qualsiasi piano orizzontale. Analogamente le equazioni di cinetica chimica in un mezzo infinitamente esteso sono invarianti rispetto al gruppo delle traslazioni, rotazioni e riflessioni.

Gli esperimenti mostrano tuttavia che fuori da queste leggi dinamiche estremamente simmetriche, si osserva la nascita di stati come una cella di Bénard o un'onda chimica, ciascuna delle quali è meno simmetrica delle leggi stesse. Questa transizione che *rompe la simmetria* scaturisce da una biforcazione. Che cosa ha ottenuto esattamente?

Per cominciare, la rottura di simmetria è una manifestazione di una *differenziazione intrinseca* fra diverse parti di un sistema, o fra il sistema e il suo ambiente. Sotto questo aspetto, essa incarna uno dei primi prerequisiti del comportamento complesso che hanno avuto luogo durante eventi come la condensazione della materia primordiale per formare galassie, o la comparsa delle prime cellule viventi. In secondo luogo, una volta assicurata una tale differenziazione, si avviano ulteriori processi che sarebbero impossibili in un mezzo non differenziato. Per esempio, come sottolineato nel capitolo I, § 7, la comparsa di eterogeneità spaziali in un mezzo biologico possono portare le cellule non differenziate di una popolazione a riconoscere il loro ambiente e a differenziarsi in cellule specializzate, permettendo cosí al materiale genetico di manifestare le sue potenzialità. Sotto questo aspetto, la rottura di simmetria appare anche come un prerequisito dell'*informazione*.

Eppure, bisogna essere consci del fatto che c'è una lunga strada prima che il prerequisito diventi condizione sufficiente. Le transizioni con rottura di simmetria, infatti, sono significative solo se portano a soluzioni asintoticamente stabili. Soprattutto, si dovrebbe essere capaci di *selezionare* in un modo o nell'altro una particolare forma asimmetrica della materia, fra le molte soluzioni che diventano possibili simultaneamente oltre il punto di biforcazione. Discuteremo ancora questi punti nei capitoli successivi.

8. *Ordine e correlazioni.*

Ci siamo ripetutamente riferiti all'equilibrio termodinamico come a uno stato in cui le caratteristiche scale spaziali sono microscopiche, e ai

fenomeni di transizione che nascono lontano dall'equilibrio come associati agli stati che godono correlazioni di raggio macroscopico. Mentre la prima affermazione appare naturale per via delle interazioni intermolecolari a corto raggio, la seconda affermazione è, al contrario, e sulle stesse basi, profondamente sorprendente. L'apparente paradosso è ancor piú esasperato se si comprende che le transizioni di non-equilibrio comprendenti, in modo coerente, una regione macroscopica, possono persino aver luogo in sistemi ideali, nei quali l'effetto delle interazioni intermolecolari può essere trascurato!

Per capire intuitivamente cosa succede, è necessario usare le fluttuazioni come sonda per investigare lo stato della materia. Supponiamo dapprima che si tratti con un sistema ideale (per esempio un gas) all'equilibrio. Siano V, N il volume totale e il numero totale di particelle, ΔV un piccolo volumetto entro V e n il numero di particelle ivi comprese. Poiché il sistema è uniforme la probabilità che ogni data particella sia nel volume ΔV è ovviamente data dal rapporto $\Delta V/V$, e la probabilità che n particelle siano simultaneamente in esso è $(\Delta V/V)^n$. Analogamente, la probabilità che una particella non si trovi in ΔV è $(V-\Delta V)/V$, e la stessa probabilità per m particelle è $[(V-\Delta V)/V]^m$. La probabilità P_n che il volume ΔV contenga n particelle è dunque

$$P_n = (\text{Numero dei modi di scegliere } n \text{ su } N \text{ particelle}) \times \left(\frac{\Delta V}{V}\right)^n \left(1 - \frac{\Delta V}{V}\right)^{N-n} \tag{2.25}$$

Nel caso interessante, ΔV è molto piú piccolo di V, e il numero n è anche piccolo paragonato a N. Il fattore combinatorio nell'equazione 2.25 si può dunque approssimare con $N^n/n!$, dove il fattoriale $n! = 1 \dots n$ viene introdotto per tener conto delle ripetizioni.

Trascurando anche n nell'esponente $N-n$, scriviamo la 2.25 come

$$P_n = \frac{1}{n!}\left(\frac{N\,\Delta V}{V}\right)^n \left(1 - \frac{\Delta V}{V}\right)^N$$

Ma $N\,\Delta V/V$ è nient'altro che il numero di particelle $\bar{n}$ nel volume ΔV. Dunque,

$$P_n = \frac{\bar{n}^n}{n!}\left(1 - \frac{\bar{n}}{N}\right)^N$$

Infine, poiché N è molto grande cosicché si può applicare la formula

$$\lim_{N\to\infty}\left(1-\frac{x}{N}\right)^N = e^{-x}$$

otteniamo

$$P_n = e^{-\bar{n}}\frac{\bar{n}^n}{n!} \tag{2.26}$$

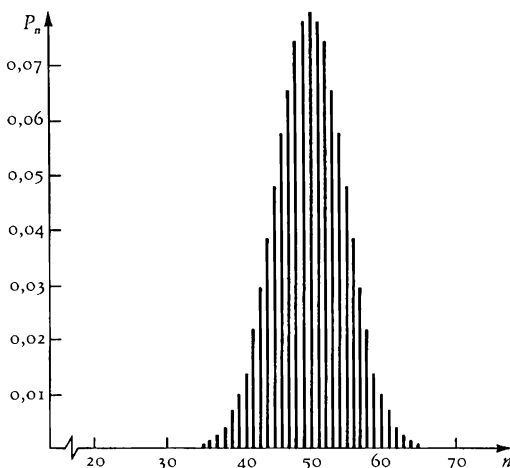
Questa è la famosa *distribuzione di Poisson*, una delle importanti distribuzioni nella teoria della probabilità. La figura 2.14 mostra che essa presenta un picco netto per un valore vicino a $\bar{n}$. Per caratterizzare questo picco, introduciamo i *momenti* $\langle n^k\rangle$ attraverso

$$\langle n^k\rangle = \sum_{n=0}^{\infty} n^k P_n$$

In particolare troviamo che: per $k = 0$ si ha

$$\sum_{n=0}^{\infty} P_n = 1\,,$$

Figura 2.14

Distribuzione poissoniana di probabilità corrispondente a $\bar{n} = 50$ in funzione di n. Notare che la distribuzione è praticamente simmetrica attorno al picco che si ha per $n = \bar{n}$.

in accordo con il significato di probabilità (condizione di normalizzazione); per $k = 1$ si ha $\langle n \rangle = \bar{n}$, come ci si aspettava. Se P_n avesse un picco infinitamente stretto attorno a $\bar{n}$ tutti i momenti conseguenti si ridurrebbero a potenze di $\bar{n}$. Inversamente, la loro deviazione da questa legge delle potenze, indicherebbe l'importanza della larghezza della distribuzione, e attraverso essa l'importanza delle fluttuazioni. Dall'equazione 2.26 si trova facilmente che

$$\langle n^2 \rangle = \bar{n}^2 + \bar{n}$$

Introducendo quindi la *varianza* di P_n

(2.27) $$\langle \delta n^2 \rangle = \langle n^2 \rangle - \langle n \rangle^2$$

troviamo che per una distribuzione poissoniana

(2.28) $$\langle \delta n^2 \rangle_p = \bar{n} = \langle n \rangle$$

Inoltre dalla derivata dell'equazione 2.26 si vede che la distribuzione congiunta della probabilità di avere n_1 e n_2 nei due sottovolumi ΔV_1, ΔV_2 (vicini o meno) è semplicemente il prodotto dei P_n individuali:

(2.29) $$P_{n_1, n_2} = P_{n_1} P_{n_2}$$

Cosí prendendo il secondo momento $\langle n_1 n_2 \rangle$ di P_{n_1, n_2} troviamo

$$\langle n_1 n_2 \rangle = \langle n_1 \rangle \langle n_2 \rangle = \bar{n}_1 \bar{n}_2$$

Introducendo la *covarianza* di P_{n_1, n_2}

(2.30) $$\langle \delta n_1, \delta n_2 \rangle = \langle n_1 n_2 \rangle - \langle n_1 \rangle \langle n_2 \rangle$$

otteniamo

(2.31) $$\langle \delta n_1 \, \delta n_2 \rangle_p = 0$$

Brevemente, siamo arrivati a due importanti quantità di natura statistica:

- la varianza, che misura l'importanza delle fluttuazioni rispetto alla media

– la covarianza, che misura il grado di correlazione fra diverse regioni spaziali nel sistema

In un regime poissoniano come quello realizzato nello stato di equilibrio di un sistema ideale, l'annullarsi della covarianza implica la totale assenza di correlazioni spaziali. Un tale stato sarà dunque il prototipo del disordine. Questo, tuttavia, non significa che le fluttuazioni siano grandi. Al contrario, l'equazione 2.28 mostra che la quantità $\sqrt{\langle \delta n^2 \rangle_p / \bar{n}^2}$, che si può considerare come una misura della loro grandezza, è uguale a $1/\sqrt{\bar{n}}$, e dunque molto piccolo se il volume ΔV è grande in confronto alla grandezza molecolare. Per esempio, se ΔV è un centimetro cubo (1 cm^3), $1/\sqrt{\bar{n}}$ sarà qualcosa come 10^{-8} per cento. Tuttavia, per ΔV di un micrometro cubo (1 μm^3) le fluttuazioni raggiungeranno valori dell'ordine di alcuni centesimi di percentuale.

Al contrario, la comparsa di una configurazione della materia caratterizzata da spazi e tempi in scala macroscopica implica che il sistema generi e mantenga relazioni riproducibili fra le sue parti distanti. Questo è descritto dalla funzione di correlazione spaziale – la covarianza $\langle \delta n(\mathbf{r}_1)\, \delta n(\mathbf{r}_2) \rangle$ – che mostrerà un'ampiezza finita e una lunghezza di correlazione macroscopica nel senso che avrà valori apprezzabili anche per distanze spaziali $|\mathbf{r}_2 - \mathbf{r}_1|$ di grandezza macroscopica. Entrambe le proprietà implicano che il sistema è stato in grado di superare la «tirannia» della poissoniana. Come ciò sia possibile, anche in situazioni nelle quali le interazioni molecolari giocano un ruolo minore, verrà discusso nel capitolo IV.

Per chiudere questo capitolo, sottolineiamo che, mentre non possiamo dare una netta definizione di complessità, cominciamo a capire alcuni dei suoi ingredienti essenziali: la comparsa di biforcazioni in condizioni lontane dall'equilibrio, e in presenza di opportune non-linearità; il verificarsi di rotture di simmetria oltre la biforcazione; e la formazione e il mantenimento di correlazioni di raggio macroscopico. Nei prossimi due capitoli tutte queste nozioni verranno gradualmente quantificate, e ne risulterà un quadro piú concreto della complessità.

Capitolo terzo

Sistemi dinamici e complessità

La complessità è una parte inseparabile del mondo dei sistemi dinamici. Henri Poincaré, il grande genio matematico vissuto a cavallo del secolo, inventò la moderna teoria dei sistemi dinamici e pose come obiettivo l'esplorare dei tipi di comportamento che possono scaturire da equazioni non-lineari accoppiate. In questo capitolo presentiamo quegli aspetti della teoria rilevanti per il problema della complessità, rimarcando in particolar modo gli ingredienti minimi dai quali nasce il comportamento complesso e le principali caratteristiche delle soluzioni delle equazioni che lo descrivono.

1. *La geometria dello spazio delle fasi.*

Già nel capitolo II, §§ 1 e 2, abbiamo introdotto le due fondamentali famiglie di sistemi dinamici che si incontrano in natura, cioè i sistemi conservativi e dissipativi. Abbiamo anche spiegato come la loro evoluzione temporale venga visualizzata attraverso il moto del punto rappresentativo nello spazio delle fasi. In questo paragrafo sviluppiamo ulteriormente questa idea esaminando i possibili tipi di moto nello spazio delle fasi. Grazie alla corrispondenza biunivoca fra di essi e il comportamento del sistema originario, otterremo in questo modo una prima classificazione qualitativa dei fenomeni che ci si può aspettare da un sistema dinamico.

Iniziamo il nostro studio con i sistemi dinamici che coinvolgono un numero finito di variabili. Questo esclude, momentaneamente, i sistemi distribuiti spazialmente descritti da variabili che dipendono con continuità dallo spazio. Questi casi verranno discussi negli ultimi due paragrafi di questo capitolo. Le equazioni di evoluzione (cfr. equazione 2.12) prendono la forma seguente

$$\frac{dX_i}{dt} = F_i(\{X_j\}, \lambda) \qquad i = 1, \ldots, n \tag{3.1}$$

nella quale non c'è chiaramente dipendenza spaziale nell'operatore F_i. Inoltre, restringiamo lo studio ai sistemi autonomi per i quali F_i non dipende esplicitamente dal tempo. Si noti che in un sistema hamiltoniano n deve essere pari, e F_i si riduce alla struttura caratteristica imposta dalla equazione 2.5*b*.

Un primo tipo di traiettoria nello spazio delle fasi compatibile con l'equazione 3.1 è data da

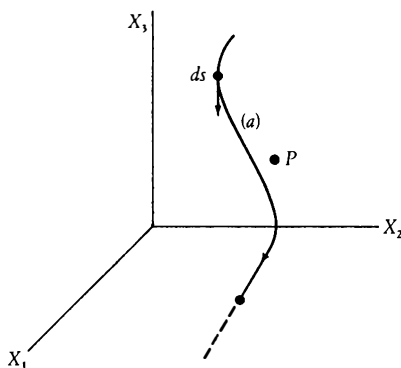
$$\frac{dX_1}{dt} = \cdots = \frac{dX_n}{dt} = 0 \tag{3.2}$$

Comprende come casi particolari gli stati di equilibrio meccanico incontrati nei sistemi conservativi e gli stati stazionari incontrati nei sistemi dissipativi. Nello spazio delle fasi tali traiettorie sono completamente degeneri, poiché sono date dalle soluzioni delle n equazioni algebriche per le n incognite $F_i = 0$. Sono dunque rappresentate da punti, come P nella figura 3.1, che chiameremo *punti fissi*.

Se le equazioni 3.2 non sono soddisfatte per ogni istante, il punto rappresentativo si muoverà lungo una traiettoria dello spazio delle fasi che definisce una curva (figura 3.1, curva a). L'elemento di linea lungo questa traiettoria per uno spostamento che corrisponde a $(dX_1, ..., dX_n)$ lungo ogni asse individuale è dato da

$$ds = \sqrt{\textstyle\sum_k dX_k^2} = \sqrt{\textstyle\sum_k F_k^2}\, dt.$$

Figura 3.1

La descrizione di un sistema dinamico nello spazio delle fasi. P: punto fisso; a: traiettoria nello spazio delle fasi.

Cosí le proiezioni della tangente alla curva lungo gli assi sono date da

$$dX_\alpha/ds = F_\alpha/\sqrt{\sum_k F_k^2}$$

e sono ben definite ovunque. Chiamiamo i punti che appartengono a tali curve *punti regolari*. Al contrario, sui punti fissi la tangente è mal definita a causa del simultaneo annullarsi di tutti gli F_i. Possiamo anche riferirci ai punti fissi come ai *punti singolari* del flusso generato dalle equazioni 3.1. L'insieme dei punti fissi e delle traiettorie dello spazio delle fasi costituirà il *ritratto delle fasi* di un sistema dinamico.

Una proprietà che gioca un ruolo decisivo nella struttura del ritratto delle fasi è il teorema di unicità delle soluzioni delle equazioni differenziali. Questo importante risultato, che risale ai lavori del matematico francese Cauchy, dice che sotto condizioni molto deboli sulla funzione F_i la soluzione corrispondente a una condizione iniziale non su un punto fisso, esiste ed è unica per tutti gli istanti in un certo intervallo $(0, T_m)$, il cui estremo superiore T_m dipende dalla specifica struttura della funzione F_i. Nella rappresentazione dello spazio delle fasi, il teorema esclude automaticamente l'intersezione di due traiettorie in ogni punto regolare.

Una seconda proprietà di grande importanza, è l'esistenza di *insiemi invarianti* del flusso e la loro struttura . Con ciò si intendono gli oggetti nello spazio delle fasi che sono limitati e rimangono immutati durante l'evoluzione generata dall'equazione 3.1. Un ovvio esempio di insieme invariante è l'insieme dei punti fissi. Un altro è la curva chiusa nello spazio delle fasi, come, per esempio, le curve *a* e *a'* della figura 2.2 che rappresentano un pendolo. Una volta che il sistema è su una tale curva, esso passa ripetutamente attraverso gli stessi stati; in altre parole, una traiettoria chiusa nello spazio delle fasi rappresenta un moto periodico.

Illustriamo la potenza delle due proprietà citate nel caso dei sistemi dinamici che evolvono in uno spazio delle fasi bidimensionale. Supponiamo che ci siano due insiemi invarianti, un unico punto fisso e una curva chiusa che circonda questo punto (figura 3.2, punto *P* e curva *a*; si noti che questo automaticamente ci limita ai sistemi dissipativi). Dal teorema di unicità, una traiettoria che parte con una condizione iniziale entro la curva *a* vi rimarrà per sempre, altrimenti dovrebbe intersecare la curva *a*. D'altra parte, le autointersezioni sono anche escluse dallo stesso teorema. Le uniche possibilità sono dunque:

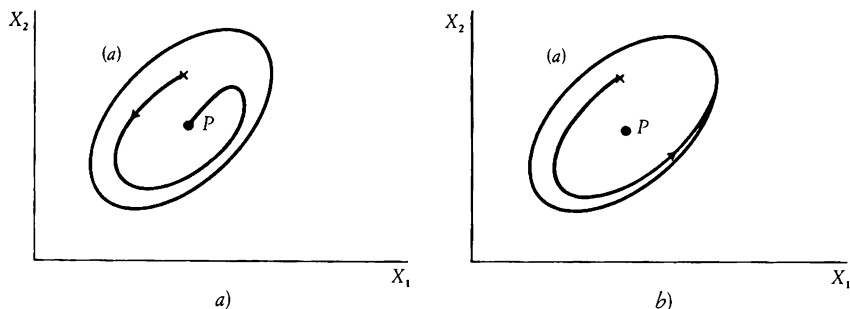
– Muoversi lungo una curva che tende al punto fisso P (figura 3.2*a*)
– Muoversi lungo una curva che tende alla curva chiusa a (figura 3.2*b*)

Vediamo quindi che l'impossibilità dell'autointersezione delle traiettorie e l'esistenza di insiemi invarianti di una certa forma determinano in larga misura la struttura del ritratto dello spazio delle fasi nello spazio delle fasi bidimensionale, e attraverso di esso il tipo di comportamento che ne può conseguire. In tre o piú dimensioni, tuttavia, i vincoli imposti da queste proprietà sono molto meno rigidi, poiché le traiettorie hanno ora molte piú possibilità di evitarsi una con l'altra «scivolando» entro il «divario» lasciato fra gli insiemi invarianti. Come vedremo piú avanti, nuovi comportamenti dinamici sono possibili da questa addizionale flessibilità, non ultimo la comparsa di evoluzioni caotiche, di tipo turbolento.

2. *Misure nello spazio delle fasi.*

Secondo la precedente discussione, l'evoluzione di un sistema dinamico si può guardare come una trasformazione dallo spazio delle fasi in se stesso. Nel corso di questa trasformazione, le traiettorie che partono da una certa parte dello spazio delle fasi possono attraversare una regione vicina, raggiungere altre parti remote di questo spazio e, forse, ivi

Figura 3.2

Ruolo degli insiemi invarianti e della non-intersezione delle traiettorie nella struttura del ritratto delle fasi. Si assume che una traiettoria chiusa a e un punto fisso P siano gli unici possibili insiemi invarianti. Ne segue che una traiettoria nello spazio delle fasi che parte da un punto interno della curva a è obbligata a seguire uno dei due moti descritti in *a*) e *b*).

restare per sempre, oppure possono tornare indietro in quella parte dello spazio delle fasi da dove sono inizialmente partite. Di regola, le diverse parti dello spazio vengono attraversate con una frequenza disuguale durante questi processi. Il problema è ora quello di caratterizzare questo complicato moto in modo che sia il piú possibile compatto e istruttivo.

Una possibilità è quella di attribuire un «peso» a ogni parte dello spazio delle fasi, e vedere quanto esso sia influenzato dall'evoluzione. Un peso naturale è il volume della parte in questione. Per vedere come l'evoluzione lo cambia, si dovrebbe considerare l'insieme di tutte le traiettorie che partono con una condizione iniziale entro questa parte, le seguono durante un intervallo di tempo t, e ricostituiscono l'insieme contenente tutti i punti rappresentativi sulle traiettorie soggetti alle sopra specificate condizioni iniziali (vedere figura 3.3). In questo modo si ottiene un volume V_t al tempo t, il cui valore iniziale viene indicato con V_0.

La figura 3.4 descrive tre casi caratteristici che dànno luogo a diversi tipi di cambiamento di V_t rispetto a V_0.

Il caso *a*, un moto circolare uniforme, è il caso piú semplice. Le equazioni del moto derivano dall'hamiltoniana $H = (X_1^2/2) + (X_2^2/2)$. Non solo si conserva il volume (area) dello spazio delle fasi, $|V_0| = |V_t|$, ma, a causa dell'isocronismo del movimento su ogni traiettoria, non è soggetto ad alcuna deformazione.

Il caso *b* descrive un moto iperbolico, le cui traiettorie hanno gli assi X_1 e X_2 come asintoti. La corrispondente hamiltoniana è $H = X_1X_2$. La

Figura 3.3

Cambiamento di un volume iniziale V_0 nello spazio delle fasi, quando i punti rappresentativi interni a questo volume evolvono sotto l'effetto della dinamica.

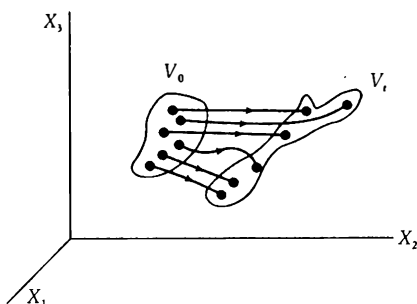
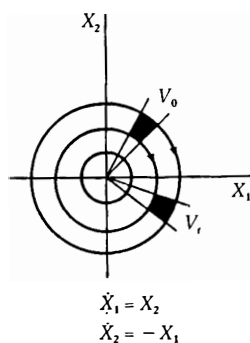
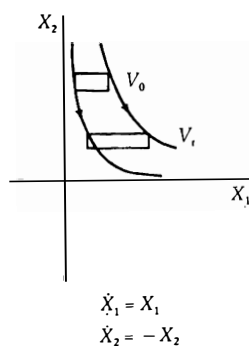
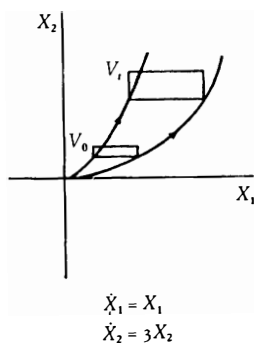



Figura 3.4
Diversi tipi di cambiamento subiti da un volume iniziale V_0 per tre rappresentativi esempi di sistemi dinamici.

Conservativo (*a*)

Conservativo (*b*)

Non conservativo (*c*)

proiezione del movimento lungo l'asse X_1 dà luogo a una crescita esponenziale, mentre lungo l'asse X_2 dà luogo a un decadimento esponenziale con lo stesso valore assoluto dell'esponente caratteristico. Per questo il volume del piano delle fasi viene ancora conservato, $|V_t| = |V_0|$, ma, contrariamente al caso precedente, è estremamente deformato. Per esempio, un rettangolo che segue il moto si espanderà lungo il suo lato parallelo a X_1 e si contrarrà lungo l'altro lato. Questo aspetto è caratteristico della classe di *moti instabili* che verranno discussi piú avanti sia in questo capitolo che nel capitolo v.

Anche il caso *c* descrive un moto instabile. Tuttavia, poiché c'è un'espansione lungo sia X_1 che X_2, il volume dello spazio delle fasi non solo viene deformato, ma aumenta anche sistematicamente. Ossia $|V_t| > |V_0|$. Chiaramente una simile evoluzione non potrebbe aver luogo indefinitamente, poiché produrrebbe alla fine un'esplosione. Se d'altro canto i segni dei membri a destra delle equazioni di evoluzione fossero opposti, si otterrebbero le stesse traiettorie, ma le frecce indicanti la direzione seguita lungo le traiettorie sarebbero invertite. In questo moto stabile il volume dello spazio delle fasi sarebbe ancora deformato, ma dovrebbe decrescere sistematicamente; ossia $|V_t| < |V_0|$. Chiamiamo quest'ultimo sistema un *sistema dissipativo*.

Mentre non si può enunciare alcuna affermazione generale concernente il cambiamento di V_t per un sistema dissipativo, per un sistema conservativo si ha che il risultato $|V_t| = |V_0|$, trovato nei due esempi della figura 3.4*a* e *b*, non è accidentale. Un teorema molto importante, dovuto al matematico francese Liouville, stabilisce infatti che il volume dello spazio delle fasi rimane invariante in *ogni* sistema conservativo.

Prima di discutere le implicazioni di questo risultato, guardiamo la deformazione del volume dello spazio delle fasi da un diverso punto di vista. In linea di principio la soluzione delle equazioni di evoluzione 3.1 costituisce un problema ben posto, nel senso che una completa specificazione dello stato $(X_1, \ldots, X_n)$ in un certo istante permette di prevedere lo stato in tutti gli altri istanti. Ma in molti casi una tale completa specificazione può essere operativamente senza significato. Per esempio, in un sistema hamiltoniano composto da particelle il cui numero è dell'ordine del numero di Avogadro, non c'è esperimento fisico che permetta di esaminare le coordinate e le velocità di tutte le particelle. E ancora, quando il moto nello spazio delle fasi diventa molto complesso, come in un regime caotico (vedi anche i §§ 7-9 di questo capitolo), non ha piú significato discutere in termini di traiettorie individuali. Sono necessari nuovi modi di descrizione, e fra questi il piú diffuso è certa-

mente una discussione del sistema in termini di concetti di probabilità. A questo scopo non consideriamo il caso isolato di un singolo sistema, ma concentriamo invece l'attenzione su un insieme di molti sistemi identici, tutti soggetti esattamente agli stessi vincoli. I membri di questo insieme, noto come *insieme di Gibbs* dallo scienziato americano che introdusse questo concetto, si trovano in generale in stati diversi. Essi si possono dunque guardare come derivanti da un insieme iniziale di sistemi i cui punti di fase rappresentativi erano contenuti in un certo volume V_0 dello spazio delle fasi, come in figura 3.3. Niente nella definizione di cui sopra limita il numero dei sistemi nell'insieme, cosí è conveniente considerare il limite in cui ogni regione, per quanto piccola, contiene un grande numero di sistemi, e richiedere la frazione di casi nell'insieme per i quali qualche variabile $\{X_i\}$ assume valori in un intervallo $\{X_1, X_1 + dX_i\}$. Nel limite continuo, questo viene espresso in termini di una *densità di probabilità*, $\rho(X_1, ..., X_n, t)$ tale che

$$\rho \, dX_1 \cdots dX_n \tag{3.3}$$

rappresenta la probabilità di trovare al tempo t un membro dell'insieme nell'elemento di volume $dX_1 \cdots dX_n$ dello spazio delle fasi che circonda $\{X_1, \cdots, X_n\}$. Richiederemo che ρ sia non-negativa e normalizzata:

$$\int \rho \, dX_1 \cdots dX_n = 1 \tag{3.4}$$

Per poter prevedere la probabilità che certe variabili assumano particolari valori nell'insieme, dobbiamo riuscire a scrivere una equazione di evoluzione per ρ. A questo fine, ragioniamo in termini molto simili alla fluidomeccanica, poiché, dopo tutto, ρ si può guardare come la densità di un fluido le cui linee di flusso seguono le traiettorie descritte dalle equazioni 3.1.

Consideriamo un elemento fisso di volume dello spazio delle fasi localizzato fra X_1 e $X_1 + dX_1$, X_2 e $X_2 + dX_2$, ecc. (figura 3.5). Al variare delle X_i secondo le 3.1, il numero dei sistemi localizzati nel volume $dX_1 dX_2 \cdots$ cambieranno nel tempo perché i sistemi entreranno e usciranno da questo volume attraverso i suoi contorni. Per esempio, i sistemi che entrano attraverso la faccia X_1 = costante al tempo t sono contenuti nel volume $\dot{X}_1 \, dt \, dX_2 \cdots dX_n = F_1 \, dt \, dX_2 \cdots dX_n$, e il loro numero è cosí dato da

$$\rho(X_1, ..., X_n, t) F_1 \, dt \, dX_2 \cdots dX_n$$

Il numero di sistemi che escono dalla faccia $X_1 + dX_1$ = costante è dato da un'espressione simile, solo che r e F_1 debbono essere valutati in $X_1 + dX_1$ invece che in X_1:

$$\rho(X_1 + dX_1, \ldots, X_n, t)F_1(X_1 + dX_1, \ldots, X_n)\, dt\, dX_2 \cdots dX_n$$
$$= \rho(X_1, \ldots, X_n, t)F_1\, dt\, dX_2 \cdots dX_n + \frac{\partial \rho}{\partial X_1} F_1\, dt\, dX_1 \cdots dX_n$$
$$+ \rho \frac{\partial F_1}{\partial X_1} dt\, dX_1 \cdots dX_n$$

dove la seconda parte dell'uguaglianza deriva da uno sviluppo in serie in un intorno di X_1 di $\rho(X_1 + dX_1)$ e di $F_1(X_1 + dX_1)$ nella quale si considerano solo i primi termini non banali.

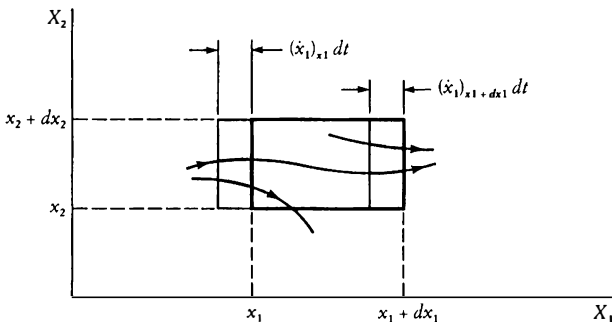
Lo stesso ragionamento si può ripetere per gli altri bordi del volume, e si arriva cosí all'equazione

(Cambiamento di ρ durante dt) $dX_1 \cdots dX_n$ = Numero di sistemi che entrano nel volume – Numero di sistemi che escono dal volume

o, dividendo per $dt\, dX_1 \cdots dX_n$ e usando le due precedenti espressioni di cui sopra

$$\frac{\partial \rho}{\partial t} = -\sum_{i=1}^{n} \left(\frac{\partial \rho}{\partial X_i} F_i + \rho \frac{\partial F_i}{\partial X_i} \right) \tag{3.5}$$

Figura 3.5

Illustrazione del bilancio del numero di sistemi di un insieme di Gibbs (area grigia) che entra in un volume dello spazio delle fasi, e successivamente lascia questo volume.

Si osservi ora che $\sum_i \partial F_i / \partial X_i$ è la divergenza del vettore $\boldsymbol{F}$, i cui componenti nello spazio delle fasi sono le variazioni $F_1 \cdots F_n$ delle singole variabili (equazione 3.1). Inoltre, se seguiamo il «fluido» delle traiettorie al variare del tempo, ricostruiamo la derivata «idrodinamica» $d\rho/dt$, nella quale appare sia l'esplicita dipendenza dal tempo di ρ sia la sua dipendenza dal tempo attraverso le variabili $X_1(t), \ldots X_n(t)$:

$$\frac{d\rho}{dt} = \frac{\partial \rho}{\partial t} + \sum_i \frac{\partial \rho}{\partial X_i} \dot{X}_i = \frac{\partial \rho}{\partial t} + \sum_i \frac{\partial \rho}{\partial X_i} F_i$$

Otteniamo cosí dall'equazione 3.5 l'importante risultato

$$\frac{d\rho}{dt} = \frac{\partial \rho}{\partial t} + \sum_{i=1}^{n} \frac{\partial \rho}{\partial X_i} F_i = -\rho \operatorname{div} \boldsymbol{F} \tag{3.6}$$

Se valutiamo div $\boldsymbol{F}$ per un sistema hamiltoniano, vediamo dall'equazione 2.5*b* che

$$\operatorname{div} \boldsymbol{F} = \sum_{i=1}^{n} \left(\frac{\partial}{\partial q_i} \dot{q} + \frac{\partial}{\partial p_i} \dot{p}_i \right) = \sum_{i=1}^{n} \left(\frac{\partial}{\partial q_i} \frac{\partial H}{\partial p_i} - \frac{\partial}{\partial p_i} \frac{\partial H}{\partial q_i} \right) = 0$$

L'equazione 3.6 implica allora che la densità di probabilità r si conserva durante il moto nello spazio delle fasi. Questo è precisamente il teorema di Liouville espresso in un linguaggio diverso, e che porta alle *equazioni di Liouville*

$$\frac{d\rho}{dt} = \frac{\partial \rho}{\partial t} + \sum_{i=1}^{n} \frac{\partial \rho}{\partial X_i} F_i = 0 \qquad \text{(sistema conservativo)} \tag{3.7}$$

Effettivamente l'equazione 3.7 rimane valida per sistemi piú generali di quelli hamiltoniani. È sufficiente per questo avere un flusso nello spazio delle fasi nel quale la densità di fase si comporta come la densità di un fluido incomprimibile, div $\boldsymbol{F} = 0$. Questa condizione si può infatti utilizzare per *definire* la classe dei sistemi conservativi. Analogamente, l'equazione 3.6 assieme a div $\boldsymbol{F} \neq 0$ è una condizione necessaria (ma non sempre sufficiente) che deve essere soddisfatta dalla classe di sistemi dissipativi.

L'equazione di Liouville per i sistemi conservativi, equazione 3.7, gioca un ruolo molto importante nella fisica statistica poiché è il punto di partenza sia per la teoria ergodica che per la teoria cinetica dei pro-

cessi irreversibili. Discuteremo questi problemi piú ampiamente nel capitolo v.

Per i sistemi dissipativi l'uso di concetti statistici, e in particolare dell'equazione 3.6, sembra essere necessaria nel caso del comportamento caotico, nel quale un moto deterministico presenta un marcato carattere casuale. Questi problemi verranno ripresi piú avanti (§ 9). Nei paragrafi successivi discuteremo nell'ordine i sistemi conservativi e dissipativi la cui dinamica è riconducibile a uno spazio delle fasi di piccole dimensioni. Nonostante la loro apparente semplicità, succede che alcuni di questi sistemi possono già presentare modi di comportamento non banali, che ricordano i fenomeni visti nei capitoli I e II.

3. *Sistemi conservativi integrabili.*

Consideriamo ora un sistema hamiltoniano descritto dalle equazioni 2.5 o 3.7. Poniamo $n = 2N$, dove N è il numero delle posizioni indipendenti o momenti del problema, spesso chiamato numero di *gradi di libertà*. Come è ben noto dall'analisi, l'integrazione di un sistema di $2N$ equazioni differenziali ordinarie equivale a trovare $2N$ integrali primi costanti del moto indipendenti, uno dei quali fornisce una relazione fra le variabili del sistema. I sistemi hamiltoniani costituiscono tuttavia una particolare classe, nella quale, sotto condizioni molto generali è sufficiente conoscere solo N integrali primi. Come Liouville dimostrò per primo, il sistema può essere integrato per semplici quadrature.

Il teorema di Liouville dice in effetti di piú. Siano I_i, per $i = 1,\dots,N$, una opportuna combinazione delle N costanti del moto. Allora il flusso hamiltoniano si può ridurre a un insieme di equazioni della forma

$$\frac{dI_i}{dt} = 0 \qquad (3.8)$$

$$\frac{d\varphi_i}{dt} = \omega_i(\{I_j\}) \qquad i = 1, \dots, N$$

Queste equazioni hanno anche una struttura hamiltoniana, grazie alla quale le ω_i sono legate all'hamiltoniana originale attraverso l'uguaglianza $\omega_i = \partial H/\partial I_i$. Chiamiamo I_i e φ_i variabili di azione e angolari rispettivamente.

Le equazioni 3.8 definiscono la classe particolare dei *sistemi integra-*

bili. Secondo il teorema di Liouville questi sistemi possono essere caratterizzati da una delle due seguenti proprietà: il possedere N integrali primi sufficientemente regolari; oppure l'essere riconducibili alle equazioni 3.8 per mezzo di appropriate trasformazioni che preservano la struttura hamiltoniana (note come trasformazioni canoniche).

Come corollario, ogni sistema hamiltoniano con un grado di libertà ($n = 2$, $N = 1$) è integrabile poiché possiede una costante del moto ($N = 1$), l'energia totale $H = E$ = costante. Il pendolo (equazione 2.6 e figura 2.2) o l'oscillatore armonico sono dunque sistemi integrabili. Un sistema hamiltoniano con due gradi di libertà è integrabile se è dato un integrale primo indipendente dall'hamiltoniana H. Per tre gradi di libertà sono necessari tre integrali primi. Nella trottola, uno dei piú frequenti problemi tridimensionali trattati nei testi introduttivi di meccanica, l'esistenza di tre integrali segue dalla conservazione dell'energia e del momento angolare. La trottola è dunque un sistema integrabile. Analogamente, questo succede anche per il problema dei due corpi in presenza di una forza centrale, la cui importanza deriva dal suo collegamento con il moto dei corpi celesti.

Piú generalmente, tutti i sistemi che possono essere separati in sistemi non accoppiati a un grado di libertà sono integrabili. I sistemi con equazioni lineari del moto sono un ovvio esempio, poiché una trasformazione di «modo normale» che diagonalizza la matrice dei coefficienti li riduce a equazioni non accoppiate a un grado di libertà. Questo è ciò che sta alla base dell'ampia letteratura sulle piccole vibrazioni attorno alla posizione di equilibrio, di cui la fisica dello stato solido nell'approssimazione armonica è un esempio particolarmente importante. Anche certi sistemi non-lineari possono portare a hamiltoniane separabili. Gli esempi piú interessanti sono i sistemi spazialmente distribuiti come il reticolo di Toda o l'equazione di Korteweg de Vries.

Studiamo ora i tipi di comportamento che ci si aspetta in un sistema integrabile. Nella rappresentazione in cui l'evoluzione è descritta in termini di costanti di moto (azioni) e i corrispondenti angoli, la risposta a questa questione è molto semplice. Per un grado di libertà si ottiene

$$\begin{aligned} I_1 &= I_{10} = \text{costanti} \\ \varphi_1 &= \omega_1 t + \varphi_{10} \end{aligned} \tag{3.9}$$

Se pensiamo a (I_1, φ_1) come alle coordinate polari, nelle corrispondenti coordinate ortogonali, P, Q, il problema diventa isomorfo all'oscillato-

re armonico. Lo spazio delle fasi è un piano, e per ogni valore di I_1 il moto è uniforme e avviene lungo la circonferenza di un cerchio (figura 3.6).

Quando vi sono molti gradi di libertà il moto è la sovrapposizione di N moti indipendenti, ciascuno descritto dall'equazione 3.9. Lo spazio delle fasi è $2N$-dimensionale, ma poiché I_1 = costante, il moto è confinato su un insieme a N dimensioni la cui proiezione nel piano delle fasi

Figura 3.6

Traiettoria nello spazio delle fasi di un sistema conservativo con un grado di libertà. Un sistema come questo è automaticamente integrabile.

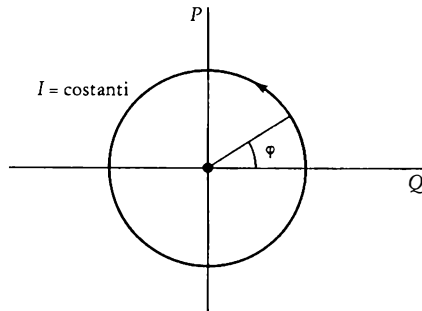
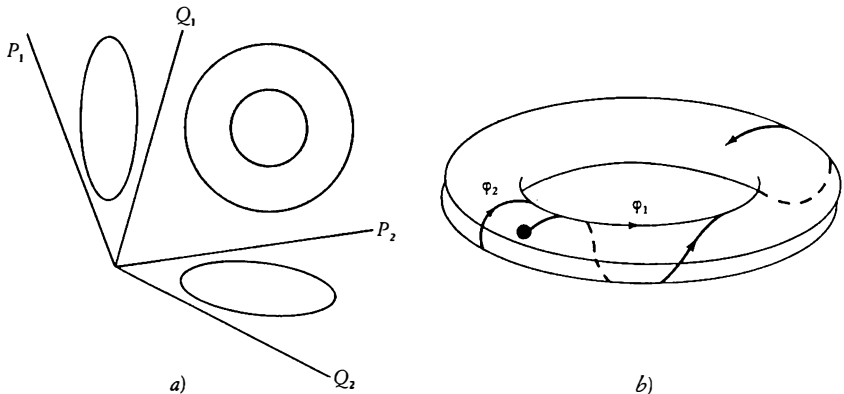



Figura 3.7

Visualizzazione delle traiettorie nello spazio delle fasi di un sistema dinamico integrabile con due gradi di libertà. Il moto ha luogo sulla superficie di un toro bidimensionale, la cui grandezza dipende da una combinazione delle variabili di azione. La conservazione dell'energia totale riduce il numero delle variabili indipendenti da quattro a tre. Il toro in *b*) è dunque inserito in uno spazio tridimensionale.

di ciascun grado di libertà è un cerchio. Un tale oggetto è chiamato *toro*, ed è visualizzato nella figura 3.7*a* e *b* nel caso $N = 2$. Notare che, a causa della conservazione dell'energia totale, lo spazio delle fasi di un sistema con due gradi di libertà è tridimensionale anziché quadridimensionale. Il toro della figura 3.7 è dunque situato in uno spazio tridimensionale.

La dipendenza dal tempo delle variabili P_i, Q_i segue immediatamente. Per $N = 1$ è periodico con periodo uguale a $2\pi/\omega_1$, mentre per $N > 1$ è la sovrapposizione di funzioni periodiche di diversi periodi. Se questi periodi sono «commensurabili», cioè se esiste un insieme di interi non nulli k_1, k_2, ... tali che

$$k_1\omega_1 + k_2\omega_2 + \cdots = 0$$

allora il risultato è chiaramente una funzione *periodica*, rappresentata da una curva chiusa.

Ma se non esistono interi non nulli per i quali l'equazione 3.10 sia soddisfatta, allora il moto è *quasi-periodico* ed è rappresentato da una spirale che si muove su un toro senza mai chiudersi su se stessa. Ciò accade, per esempio, per $N = 2$ se $\omega_1 = 1$ e ω_2 è un numero irrazionale, ad esempio $\omega_2 = \sqrt{2}$.

Un moto quasi-periodico può risultare assai complicato. Poiché la sua traiettoria rappresentativa non si chiude mai e non interseca mai se stessa in virtú del teorema di unicità discusso nel § 1, «riempie» gradualmente il toro nel senso che è infine obbligata a passare attraverso un prescritto intorno di ogni punto della superficie toroidale. Questi moti sono detti «densi ovunque».

Come ripetutamente sottolineato nel capitolo II, specialmente al § 6, un sistema conservativo non può godere della stabilità asintotica. Dunque né i moti periodici, né quelli quasi-periodici sopra descritti si possono usare come modelli di attrattori. È facile comprenderne il perché: dalla prima relazione 3.9 esiste una continuità di valori possibili per la costante del moto I_i. Cosí, invece delle figure 3.6 e 3.7*b*, abbiamo in effetti un ritratto delle fasi che comprende un continuum di orbite chiuse o di superfici toroidali contenute una dentro l'altra. Inevitabilmente, una piccola perturbazione che rimuove il sistema da una data orbita lo getterà su un'altra superficie invariante, sulla quale il sistema rimarrà intrappolato finché un'altra perturbazione lo manderà su di una ancora diversa superficie invariante, e cosí via. Al piú ci aspettiamo dunque una stabilità semplice o orbitale e un moto nervoso, anziché un

comportamento perfettamente riproducibile sotto l'effetto delle perturbazioni.

I moti in cui un continuum di insiemi chiusi invarianti circonda un punto di equilibrio, sono detti *ellittici*, e lo stesso punto di equilibrio è detto *centro*. Secondo la discussione del capitolo III, § 2 (vedere soprattutto la figura 3.4*a*), ci aspettiamo che questi moti diano origine a deformazioni del volume dello spazio delle fasi piuttosto modeste (che naturalmente sono conservati per il teorema di Liouville).

Fino ad ora abbiamo descritto il comportamento dei sistemi integrabili nella rappresentazione delle variabili azione-angolo. Come spiegato all'inizio di questo paragrafo, queste variabili sono collegate a quelle originali da una trasformazione. In qualche caso eccezionale questa trasformazione è piuttosto semplice. Per esempio, nel problema dell'oscillatore armonico

$$H = \frac{1}{2} m\dot{x}^2 + k\frac{x^2}{2}$$

I e φ sono dati da

$$x = \left(\frac{2I}{\sqrt{mk}}\right)^{1/2} \sin\varphi,$$

$$\dot{x} = \left(2I\sqrt{\frac{k}{m^3}}\right)^{1/2} \cos\varphi$$

e l'hamiltoniana si riduce a

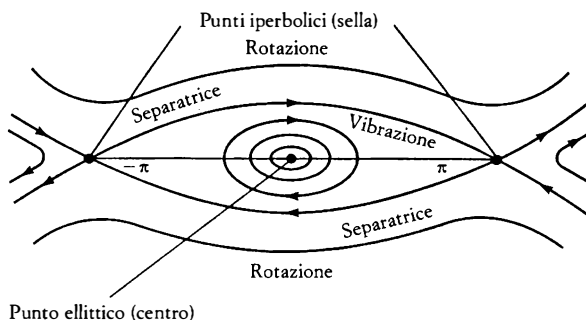
$$H = \sqrt{\frac{k}{m}}\, I = \omega I$$

dove ω è la frequenza angolare dell'oscillatore. In generale, tuttavia, la trasformazione è fortemente non-lineare. Ci aspettiamo quindi che le traiettorie siano versioni fortemente deformate delle figure illustrate nella figura 3.6 e 3.7*b*. Inoltre, certi tipi di moto che avvengono nello spazio delle fasi originario diventeranno molto artificiali nella rappresentazione azione-angolo per la semplice ragione che la variabile «angolo» cesserà di essere un «angolo» che soddisfa la condizione $0 \leqslant \varphi \leqslant 2\pi$. Questo è il caso dei moti instabili, come il moto iperbolico mostrato in figura 3.4*b*, in cui alcune delle variabili diventano non limitate, e, in concomitanza, i periodi di certe traiettorie tendono all'infinito. Nel pendolo (cfr. figura 2.2) questo avviene vicino al punto di equilibrio in-

stabile $\theta = \pm\pi$. La regione delle oscillazioni limitate (vibrazioni) è separata dalla regione del moto illimitato (rotazione) dal particolare tipo di traiettoria nota come *separatrice* (figura 3.8), che congiunge gli equilibri instabili $\theta = \pi$ e $\theta = -\pi$. Poiché la derivata di una variabile tende a zero vicino a un punto fisso, il tempo necessario per muoversi da un punto della separatrice al punto fisso è infinito. La configurazione peculiare delle traiettorie vicino alla separatrice, che assomiglia alle superfici di livello attorno a un passo di montagna, è all'origine del termine *punto di sella* usato per caratterizzare gli equilibri instabili nei sistemi conservativi. Come già suggerito nella figura 3.4, il moto vicino a un punto di sella dà luogo a grandi deformazioni del volume dello spazio delle fasi, mentre conserva il suo valore numerico.

In breve, a parte forse le separatrici e i punti di sella che possiedono, in forma nascosta, ciò che risulterà essere un principale meccanismo del comportamento complesso, si può dire che il mondo dei sistemi integrabili è un mondo semplice e tranquillo. L'attento lettore penserà che non è cosí sorprendente, poiché in un sistema integrabile ci si libera effettivamente delle interazioni e si tratta con entità indipendenti. Ma questa dovrebbe essere una situazione patologica, come vedremo piú ampiamente in un successivo paragrafo. È notevole comunque che, nonostante il loro carattere eccezionale, i sistemi integrabili dominarono la scienza fino agli anni '50, e costituiscono tuttora i soggetti principali della

Figura 3.8

Ritratto delle fasi di un pendolo. Attorno al punto fisso ellittico, che descrive l'unica posizione di equilibrio stabile, si può verificare un continuum di moti periodici vibrazionali la cui ampiezza e il cui periodo dipendono dalle condizioni iniziali. I punti fissi iperbolici a $\pm\pi$ descrivono equilibri instabili. Essi sono uniti da un cappio chiuso formato da due segmenti di separatrici. Una condizione iniziale al di fuori della curva dà luogo a un moto rotazionale libero del pendolo.

maggior parte dei testi di meccanica. Il loro grande ruolo storico e il loro indiscutibile valore pedagogico sono certamente spiegazioni parziali di questa situazione paradossale. Ma nel concludere questo paragrafo è necessario rimarcare che per i nostri scopi il mondo dei sistemi integrabili non può essere il luogo di nascita della complessità. Dobbiamo guardare altrove, e questa ricerca sarà l'obiettivo principale dei paragrafi seguenti di questo capitolo.

4. *Biforcazione in semplici sistemi dissipativi: ricerca di archetipi della complessità.*

Volgiamo ora la nostra attenzione ai sistemi dissipativi. Sceglieremo un semplice modello matematico astratto, che ci aiuterà a realizzare alcuni dei meccanismi che prendono parte al cambiamento qualitativo del comportamento, e piú tardi lo collegheremo al mondo reale dei sistemi fisico-chimici.

Consideriamo una singola variabile x che evolve secondo la

$$\frac{dx}{dt} = f(x, \lambda) = -x^3 + \lambda x \tag{3.10}$$

Questo semplice sistema è controllato da un singolo parametro λ, dal quale la funzione derivata f dipende in un modo perfettamente regolare. Determiniamo dapprima i punti fissi (stati stazionari):

$$-x_s^3 + \lambda x_s = 0$$

Una soluzione di questa semplice equazione algebrica che esiste sotto ogni condizione è quella banale

$$x_0 = 0 \tag{3.11a}$$

Possono tuttavia esistere altre soluzioni non banali. Esse soddisfano l'equazione che si ottiene eliminando x_s dalla precedente equazione cubica:

$$-x_s^2 + \lambda = 0$$

Se λ è negativo questa equazione è senza significato poiché dà una solu-

zione complessa, che non può descrivere una situazione fisica. Ma se λ è positivo l'equazione ammette la coppia di soluzioni:

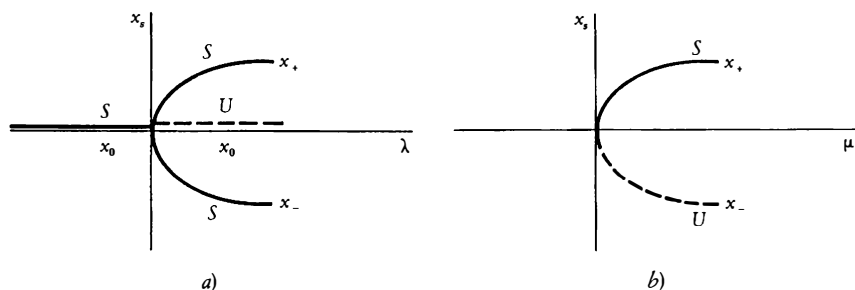
$$x_{\pm} = \pm\sqrt{\lambda} \qquad (3.11b)$$

che si unisce con x_0 per $\lambda = 0$ ma si distacca da esso per $\lambda > 0$. Questo è il fenomeno della *biforcazione* definito nel capitolo II, § 7, e illustrato nella figura 2.12; è conosciuto anche come *biforcazione a forcone*. Nella figura 3.9*a* si è disegnato un simile diagramma, adattato alla struttura del problema specifico. Le linee continue e tratteggiate sono state utilizzate per indicare la stabilità asintotica e le soluzioni instabili rispettivamente. Per il semplice modello considerato, il problema della stabilità è banale, poiché l'equazione 3.10 si può risolvere esattamente con metodi standard. Si trova che x_0 è globalmente asintoticamente stabile per $\lambda < 0$ e instabile per $\lambda > 0$, e che x_+, x_- sono asintoticamente (ma non globalmente) stabili. In altre parole i rami $x_{\pm}$ biforcano nella direzione nella quale lo «stato di riferimento» $x_0 = 0$ perde la sua stabilità e sono essi stessi stabili. Chiameremo questo *biforcazione supercritica*. Invero, la relazione fra biforcazione supercritica e stabilità non è casuale. Un risultato generale della teoria della biforcazione, discusso in dettaglio nell'*Appendice* B, stabilisce che sotto certe condizioni sulla funzione derivata f, spesso soddisfatta nelle applicazioni pratiche, i rami supercritici sono stabili, e quelli sottocritici sono instabili.

Figura 3.9

Due esempi di biforcazione elementare che avvengono in un sistema dinamico dissipativo.

a) Biforcazione simmetrica (forcone), in cui uno stato di riferimento x_0 perde la sua stabilità per $\lambda > 0$ e dà luogo a due rami di soluzioni x_+, x_-, che si fondono in x_0 per $\lambda = 0$.

b) Biforcazione a punto limite, dove un ramo stabile (x_+) ed uno instabile (x_-) collidono e si annullano uno con l'altro per $\mu = 0$.

Ma il risultato piú sorprendente del nostro semplice modello è certamente illustrato dall'equazione 3.11*b*. Ricordiamo che avevamo in origine una semplice equazione cubica che dipende regolarmente dal parametro λ. Ci accorgiamo ora che questa dinamica molto regolare ha dato luogo a una *singolarità*. Infatti, nelle vicinanze del punto di biforcazione $\lambda = 0$, le soluzioni $x_{\pm}$ non si possono espandere in serie di potenze in funzione del parametro. Dipendono dunque in modo *non-analitico* da λ. Questa è la ripercussione, a livello matematico, del cambiamento qualitativo di comportamento introdotto dal fenomeno della biforcazione.

Prendiamo un altro esempio ancora piú semplice. Consideriamo il sistema dissipativo

$$\frac{dx}{dt} = -x^2 + \mu \tag{3.12}$$

dove il parametro di controllo è ora indicato con μ. I punti fissi sono dati da

$$x_{\pm} = \pm\sqrt{\mu} \tag{3.13}$$

che sono messi su grafico in funzione di μ nella figura 3.9*b*.

L'equazione dipendente dal tempo si può risolvere ancora esattamente, e le soluzioni mostrano che il ramo x_- è instabile, mentre il ramo x_+ è asintoticamente stabile. Notiamo che, al decrescere di μ dai valori positivi, i rami stabile e instabile «collidono» per $\mu = 0$ e successivamente si annullano. Per questa ragione chiamiamo $\mu = 0$ un *punto limite*, o piega. La sua presenza segnala ancora una volta la comparsa di singolarità, come illustrato dall'equazione 3.13.

Combiniamo ora i due esempi precedenti considerando il sistema dissipativo

$$\frac{dx}{dt} = -x^3 + \lambda x + \mu \tag{3.14}$$

(non aggiungiamo il termine quadratico dell'equazione 3.12 poiché lo si potrebbe eliminare con un opportuno cambiamento di variabili). Abbiamo in questo modo un problema che coinvolge i due parametri λ e μ. I punti fissi sono dati dalla equazione cubica canonica

$$-x_s^3 + \lambda x_s + \mu = 0$$

Dall'algebra elementare sappiamo che questa equazione può avere fino a tre soluzioni. Inoltre, al variare dei parametri le tre soluzioni si fondono e dunque si rimane con una sola soluzione. Si possono definire regioni nello spazio dei parametri che separano questi due regimi, precisamente

$$-4\lambda^3 + 27\mu^2 = 0 \qquad (3.15)$$

Questa curva è rappresentata nella figura 3.10*a*. La regione delle tre soluzioni reali finisce in un punto (l'origine 0 nella figura), nel quale c'è una dipendenza singolare di λ da μ. Questa è nota come *singolarità cuspidale*.

Le figure 3.10*b* e *c* forniscono due diverse prospettive della dipendenza delle soluzioni dai parametri. Nella 3.10*b* tracciamo x_s in funzione di μ per un fissato λ. Otteniamo una curva a esse che indica la coesistenza di soluzioni multiple per dati valori dei parametri. Inoltre due di questi rami sono simultaneamente stabili. La regione di bistabilità termina ai due punti limite μ_1 e μ_2 in vicinanza dei quali si osserva il comportamento mostrato nella figura 3.9*b*. Ci si può facilmente convincere che sotto queste condizioni un aumento di μ oltre μ_1 e fino a μ_2, seguito da un decremento nella direzione opposta, porterà a un *ciclo di isteresi*, un fenomeno già introdotto nel capitolo 1, § 4, in connessione con gli esperimenti sulla reazione di Belusov-Žabotinskij.

La figura 3.10*c* mostra x_s in funzione di λ per un fissato μ. Otteniamo in questo caso due curve disgiunte, una (*a*) definita per tutti i valori di λ, e un'altra (*b*), definita solo per $\lambda \geq \bar{\lambda}$, punto in cui mostra una singolarità a piega. Per $\lambda < \bar{\lambda}$ è possibile solo una soluzione stabile, mentre per $\lambda > \bar{\lambda}$ si ha bistabilità come prima. Infine, la figura 3.10*d* combina le informazioni contenute nella figura 3.10*a-c* grazie a un diagramma tridimensionale delle soluzioni in funzione dei parametri λ e μ.

È importante osservare che per nessun valore di $|\mu|$ non nullo, per quanto piccolo, si può osservare la biforcazione simmetrica illustrata nella figura 3.9*a*. Dunque μ agisce come un'*imperfezione* che distrugge la biforcazione. D'altra parte, il punto limite, figura 3.9*b*, si mostra essere robusto, nel senso che si ritrova in entrambe le figure 3.10*b* e *c*. Ma se entrambe λ e μ vengono variate simultaneamente, ci sarà anche una particolare combinazione di valori ($\mu = 0$, λ variabile nel nostro caso) per i quali si riottiene la biforcazione, quando il sistema sarà in grado di attraversare la singolarità cuspidale (figura 3.10*a*) in modo simmetrico.

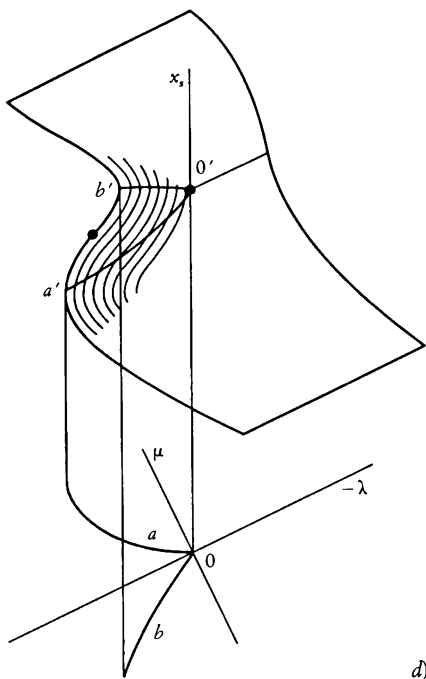
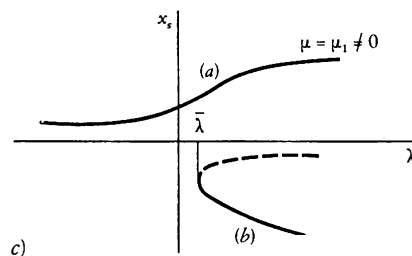
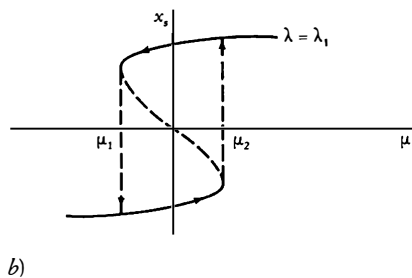
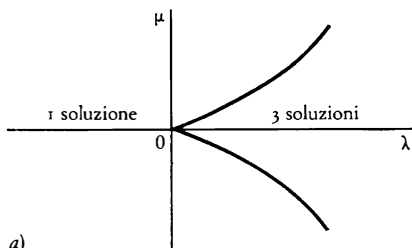
Figura 3.10

Effetto dei parametri nella biforcazione delle soluzioni a stato stazionario dell'equazione 3.14.

a) Curva dello spazio dei parametri che delimita la regione di esistenza di tre soluzioni reali.

b) Isteresi della soluzione per un fissato λ, al variare di μ. Il punto limite di biforcazione rimane robusto.

c) Distruzione della biforcazione a forcone quando il parametro μ, che agisce come imperfezione, non è identicamente zero.

d) Veduta tridimensionale che combina le informazioni di *a*), *b*) e *c*).

La discussione precedente illustra un profondo concetto di grande importanza nella teoria dei sistemi dinamici, ovvero la *stabilità strutturale*. Essa mostra infatti che certi fenomeni come la biforcazione a forcone avvengono solo se i parametri presenti soddisfano almeno una uguaglianza. Poiché in un sistema fisico una tale stretta uguaglianza è difficile da realizzare ci aspettiamo che questi fenomeni scompariranno sotto deboli cambiamenti dei valori dei parametri. Questi sono chiamati *strutturalmente instabili*. D'altra parte esistono altri fenomeni, come la biforcazione a punto limite (figura 3.9*b*), che sussistono (sebbene essi possano essere traslati) sotto cambiamenti dei parametri di controllo che influiscono sulla struttura delle leggi di evoluzione. Questi sono chiamati *strutturalmente stabili*. Si noti che secondo questa definizione *tutti i sistemi conservativi sono strutturalmente instabili*, poiché la presenza di piccoli termini dissipativi (come un debole attrito in un pendolo) alterano qualitativamente il ritratto delle fasi e conferiscono la proprietà della stabilità asintotica a certe soluzioni privilegiate, gli attrattori.

Dal punto di vista delle applicazioni fisiche, ciò che abbiamo ottenuto in questo paragrafo, è stato di identificare qualche modello matematico «minimale» che dia origine a uno dei fenomeni di transizione osservati sperimentalmente che accompagnano il comportamento complesso, e cioè la comparsa di soluzioni multiple simultaneamente stabili che descrivono la capacità di avviare ed eseguire compiti regolatori. Dobbiamo ancora convincerci che tali modelli matematici hanno interessanti realizzazioni fisiche. Posticipiamo questo compito al capitolo III, § 6. Chiudiamo la discussione aggiungendo che nei sistemi a una variabile che contengono non-linearità di ordine piú alto sono possibili fenomeni di transizione piú complessi. La teoria delle catastrofi, uno speciale ramo dei sistemi dinamici, punta alla classificazione di questi fenomeni e, in particolare, alla determinazione delle relazioni fra i parametri che delimitano qualitativamente diversi tipi di comportamento.

5. *Sistemi dissipativi nello spazio delle fasi bidimensionale-cicli limite.*

L'evoluzione dei sistemi a una variabile considerati nel paragrafo precedente avvenivano nello spazio delle fasi a una dimensione. Nella terminologia usata nel § 1, gli unici insiemi invarianti che possono esistere in un tale spazio sono i punti fissi (stati stazionari). Le traiettorie di fase sono semplicemente semirette che convergono o divergono dai punti fissi, a seconda delle loro proprietà di stabilità, come illustrato

nella figura 3.11*a*, *b* per i modelli descritti dalle equazioni 3.10 e 3.12 rispettivamente.

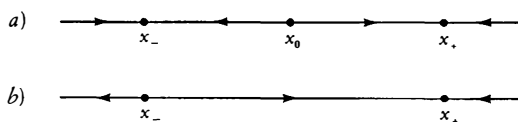
In uno spazio delle fasi bidimensionale, considerazioni topologiche suggeriscono chiaramente una flessibilità considerevolmente piú grande, ma la nostra discussione del § 1 mostra che rimangono condizioni molto rigide che limitano le varie possibilità. Un sistema a due variabili evolve in un tale spazio delle fasi. Naturalmente ci aspettiamo che esso possegga gli insiemi invarianti dei sistemi a una variabile, e cioè i punti fissi. Inoltre, poiché i sistemi conservativi costituiscono una classe particolare di sistemi a due variabili, ci aspettiamo di incontrare, sotto certe condizioni, il ritratto delle fasi dei sistemi integrabili discussi nel § 3.

In particolare i punti fissi possono comportarsi come selle o centri; può esistere un continuo di traiettorie chiuse che circondano il centro, e le separatrici dei punti di sella possono curvarsi a formare cappi che uniscono due stati instabili come nella figura 3.8. Il problema è se, in aggiunta a ciò quanto sopra, ci si possano aspettare nuove possibilità collegate specificatamente al carattere dissipativo del sistema, grazie alla ulteriore dimensione dello spazio delle fasi.

La figura 3.12 descrive una prima possibilità, collegata al modo in cui il punto fisso P viene avvicinato nel tempo. In entrambi i casi illustrati, P è asintoticamente stabile, ma le descrizioni delle fasi nelle sue vicinanze sono completamente differenti (questo era certamente impossibile nella figura 3.11). In (*a*), le traiettorie tendono a P quando il tempo tende all'infinito in modo piú o meno radiale, ciò che corrisponde a una caduta esponenziale nel tempo della perturbazione attorno a P. Chiamiamo questa configurazione un *nodo* (stabile). In (*b*), d'altra parte, le traiettorie si muovono a spirale attorno a P finché, alla fine, convergono verso di esso. Se le variabili dipendono dal tempo questo si manifesterà sotto forma di oscillazioni smorzate. Chiamiamo questa configurazione un *fuoco* (stabile).

Figura 3.11

Traiettorie nello spazio delle fasi di un sistema dinamico dissipativo a una variabile.

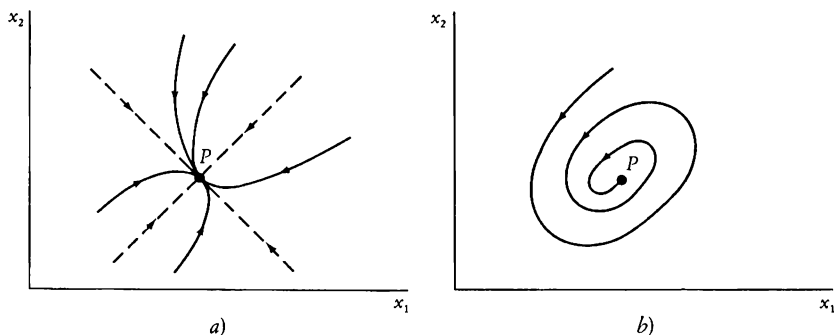
a) Una soluzione instabile e due asintoticamente stabili come nella figura 3.9*a*.
b) Una soluzione instabile e una asintoticamente stabile come in figura 3.9*b*.
Le condizioni iniziali alla sinistra di x portano a un comportamento esplosivo.

Si noti che è possibile anche la configurazione in cui tutte le frecce sono invertite, ma in questo caso si tratta di un nodo o di un fuoco instabili. Qualunque sia la loro stabilità, i nodi o i fuochi possono esistere solo nei sistemi dissipativi, poiché essi violano chiaramente l'invarianza rispetto all'inversione temporale.

Ma la possibilità piú interessante è certamente quella descritta dalla figura 3.13. Nel campo dello spazio delle fasi considerato, il sistema possiede un singolo punto fisso, P, che per i valori dei parametri qui considerati, è instabile. Le traiettorie che partono vicino a P non possono essere attratte da esso e tendono a divergere. Ma in una situazione fisicamente ragionevole la divergenza non può portare all'infinito poiché questa implicherebbe che il sistema può assorbire o cedere una infinita quantità di energia e materia. Ci aspettiamo dunque che le traiettorie rimangano confinate e che soddisfino ancora le richieste esaminate nel § 1 (niente autointersezione, ecc.). In uno spazio bidimensionale l'unica possibilità è che le traiettorie finiscano in una curva chiusa, indicata da C nella figura 3.13. Chiamiamo queste curve, scoperte da Poincaré, *cicli limite*. Nel particolare esempio scelto nella figura 3.13, il ciclo limite è asintoticamente stabile. Ma una volta che constata la possibilità di esistenza di curve chiuse che si trovano a una distanza finita da ogni altra curva chiusa (contrariamente ai sistemi conservativi), si

Figura 3.12

Possibile configurazione dello spazio delle fasi attorno ad un punto fisso P di un sistema dinamico dissipativo che coinvolge due variabili, in aggiunta alle configurazioni (sella e centro) permesse ai sistemi conservativi.

a) Nodo asintoticamente stabile. Le soluzioni evolvono monotonamente verso P.

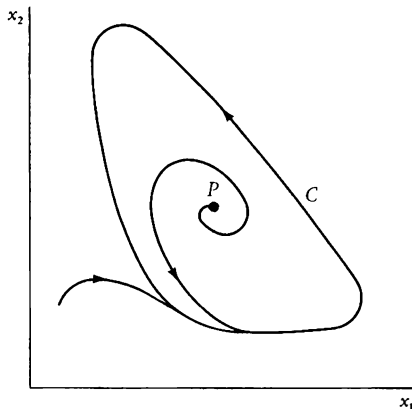
b) Fuoco asintoticamente stabile. Le soluzioni evolvono verso P mentre eseguono oscillazioni smorzate. In entrambi i casi P è un attrattore zero dimensionale (punto).

possono intravedere altre configurazioni. Per esempio, C può essere orbitalmente instabile, e le traiettorie respinte possono tendere a un altro ciclo limite o a un punto fisso stabile. In ogni caso, secondo quanto detto nel § 1, una volta su un ciclo limite, il sistema seguirà la stessa successione di stati per sempre; in altre parole, mostrerà un comportamento periodico nel tempo. Un ciclo limite asintoticamente stabile costituisce un *attrattore periodico*. Grazie alla sua stabilità asintotica, esso sarà robusto nei confronti delle perturbazioni e costituirà cosí il naturale archetipo per descrivere i fenomeni ritmici riproducibili osservati in natura, come visto nel capitolo 1.

Studiamo ora i meccanismi che presiedono alla formazione di un ciclo limite. Per avere una chiara idea delle richieste minimali e per evitare tecnicismi il piú possibile, consideriamo dapprima un modello esattamente risolubile. Immaginiamo un sistema conservativo a due variabili (automaticamente integrabile secondo il § 3), come, ad esempio, un oscillatore armonico. Nelle variabili d'azione e d'angolo l'evoluzione è data da un insieme di equazioni simile a 3.8:

$$\frac{dr}{dt} = 0$$

$$\frac{d\varphi}{dt} = \omega$$

Figura 3.13

Attrattore periodico: le traiettorie nello spazio delle fasi evolvono verso un ciclo limite C, che circonda un punto fisso instabile P.

in cui la variabile d'azione è ora indicata da r per ragioni che diverranno piú chiare fra breve. Per un dato valore di r questo sistema segue una traiettoria chiusa come in figura 3.6, che ricorda un ciclo limite, ma che ne differisce tuttavia per il fatto che è circondato da un continuo di altre traiettorie. Modifichiamo dunque il succitato sistema con una *perturbazione dissipativa*, che viola la simmetria dell'inversione temporale: la scegliamo in modo tale che possa ancora esistere un punto fisso e che il risultante ciclo limite sia asintoticamente stabile almeno sotto certe condizioni. Chiaramente non ha certo molto senso applicare una perturbazione dissipativa alla variabile di fase φ, che aumenta «passivamente» da 0 a 2π a ogni giro della traiettoria periodica. Perturbiamo pertanto la parte radiale aggiungendo un membro non banale a destra. Secondo il § 4 (cfr. equazione 3.10), la perturbazione piú semplice che soddisfa queste richieste è una cubica. Scriviamo dunque:

$$\frac{dr}{dt} = \lambda r - r^3 \qquad \frac{dr}{dt} = \omega \tag{3.16}$$

La prima equazione è identica alla 3.10, con l'ulteriore richiesta che il raggio r sia positivo. Il comportamento delle sue soluzioni segue immediatamente dalla figura 3.9*a* ed è mostrata nella figura 3.14. Per $\lambda<0$ esiste solo la soluzione banale $r=0$ ed è stabile. Nello spazio delle fasi definito dalle coordinate ortogonali

$$x_1 = r\cos\varphi, \qquad x_2 = r\sin\varphi$$

le traiettorie convergeranno al punto fisso $x_1 = x_2 = 0$, che si comporterà come un fuoco in vista del fatto che la variabile di fase φ varia secondo la

$$\varphi = \varphi_0 + \omega t \tag{3.17}$$

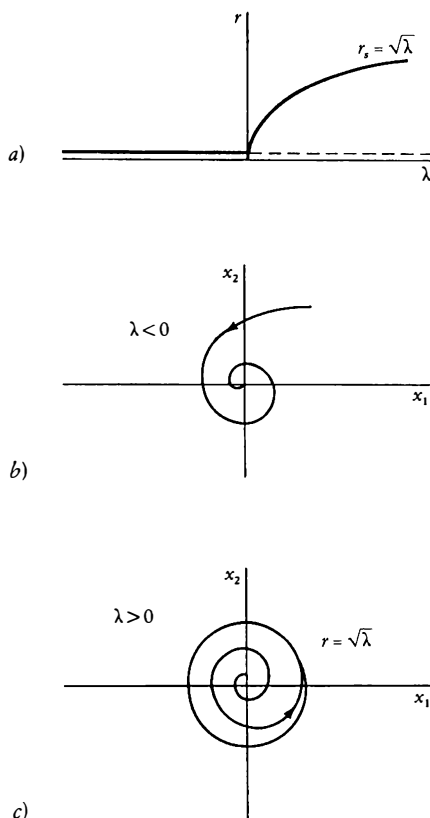
φ_0 essendo la fase iniziale (cfr. figura 3.14*b*). Ma quando $\lambda>0$ la prima equazione 3.16 ammette la soluzione non banale $r_s = \sqrt{\lambda}$, che è asintoticamente stabile, mentre il punto fisso $r=0$ è instabile. Nello spazio delle fasi (figura 3.14*c*) questo viene rappresentato dalle traiettorie che partono lungo una spirale dal punto fisso e che convergono alla curva chiusa $r_s = \sqrt{\lambda}$ – il ciclo limite. Dunque $\lambda = 0$ è un punto di biforcazione di tipo nuovo, noto come *biforcazione di Hopf*. Contrariamente alle bi-

forcazioni statiche incontrate nel § 4, che alla fine portavano tutte a soluzioni indipendenti dal tempo, una biforcazione di Hopf è un fenomeno dinamico che porta a un comportamento periodico nel tempo. L'ampiezza della soluzione cresce in modo regolare dallo zero con l'allontanarsi del parametro λ dal suo valore nel punto di biforcazione, come si vede chiaramente nella figura 3.14*a*. Inoltre, essa dipende dalla distanza dalla biforcazione in modo non analitico. D'altra parte, il periodo

Figura 3.14

Biforcazione di Hopf illustrata in base al semplice modello delle equazioni 3.16.

a) Raggio di un ciclo limite in funzione del parametro di controllo.

b) e *c*) Ritratti delle fasi, rispettivamente per $\lambda < 0$ e per $\lambda > 0$.

dell'oscillazione è generalmente finito già al punto di biforcazione, fintanto che ω non è identicamente uguale a zero nell'equazione 3.16. In ogni caso, l'osservazione importante è che sia il periodo che l'ampiezza sono intrinseci al sistema poiché essi dipendono solo dai valori di λ e ω che sono costruiti nell'equazione di evoluzione. Questo è sorprendentemente differente dai sistemi conservativi nei quali le caratteristiche delle oscillazioni dipendono dalle condizioni iniziali. Abbiamo cosí un semplice modello matematico che riproduce gli aspetti essenziali di una grande classe di fenomeni ritmici fra quelli incontrati nel capitolo 1.

Dal momento che r e φ hanno proprietà rispettivamente di una variabile radiale e angolare, si può mostrare che queste equazioni non risentono della sensibilità alle perturbazioni che caratterizzavano le soluzioni dell'equazione 3.10. Cosí *una biforcazione di Hopf è un fenomeno strutturalmente stabile*.

I sistemi dissipativi a due variabili possono dar luogo a fenomeni di transizione piú intricati. Le corrispondenti biforcazioni sono, tuttavia, strutturalmente instabili nel senso che esse richiedono che sia soddisfatta una uguaglianza fra i parametri. Riguardo questo aspetto vale la pena notare che la teoria dei sistemi dinamici è riuscita a classificare completamente tutte le descrizioni delle fasi strutturalmente stabili che possono crearsi in sistemi che evolvono in uno spazio delle fasi bidimensionale.

6. *Riduzione a sistemi a poche dimensioni: parametri d'ordine e forme normali.*

Negli ultimi due paragrafi abbiamo visto che certi fenomeni che avvengono nello spazio delle fasi a poche dimensioni sussistono quando la struttura delle equazioni di evoluzione viene modificata attraverso un cambiamento di parametri. Ci si deve dunque aspettare che questi fenomeni siano *tipici* almeno entro il dominio dei sistemi a una o due variabili, nel senso che essi dovrebbero essere osservabili per grandi classi di questi sistemi entro certi intervalli dei valori dei parametri. Convinciamoci di questo considerando pochi semplici modelli che non sono piú astratti esempi matematici, ma che derivano invece dalle leggi della fisica e della chimica.

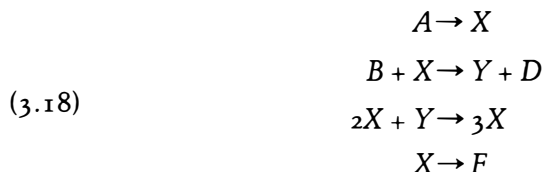
Biforcazioni a forcone e a punto limite.

Questi fenomeni sono ben illustrati da modelli derivati dalla cinetica chimica. Come discusso nei capitoli I, § 4, e II, § 4, una reazione chimica implica di regola collisioni fra almeno due molecole, la cui frequenza è con buona approssimazione proporzionale al prodotto delle concentrazioni delle specie considerate. Questo genera immediatamente non-linearità polinomiali, che erano alla base dei risultati derivati nel § 3.

Una utile visualizzazione di queste possibilità è fornita dal modello autocatalitico a una variabile introdotto nell'equazione 2.15. Come mostrato nell'equazione 2.17 i punti fissi sono le soluzioni di un'equazione cubica. All'equilibrio questa equazione ammette una sola soluzione, ma lontano dall'equilibrio il cambiamento alternante dei segni dei vari termini suggerisce fino a tre soluzioni reali. Ora, come è ben noto dall'algebra, ogni tale cubica si può trasformare in una «forma canonica» equivalente all'equazione 3.14 con un semplice cambiamento lineare di variabili. Ne segue che il modello 2.15, compatibile con le richieste imposte dalla termodinamica e dalla cinetica chimica, dà origine a tutti i fenomeni statici di biforcazione discussi nel § 4.

Biforcazione di Hopf e cicli limite.

Anche qui la cinetica chimica ci procura degli esempi chiarificatori. Uno degli esempi piú ampiamente studiato è il modello trimolecolare o *brussellatore*. Esso implica due intermediari accoppiati e un passo autocatalitico, e per semplicità è generalmente considerato nel limite delle reazioni irreversibili:

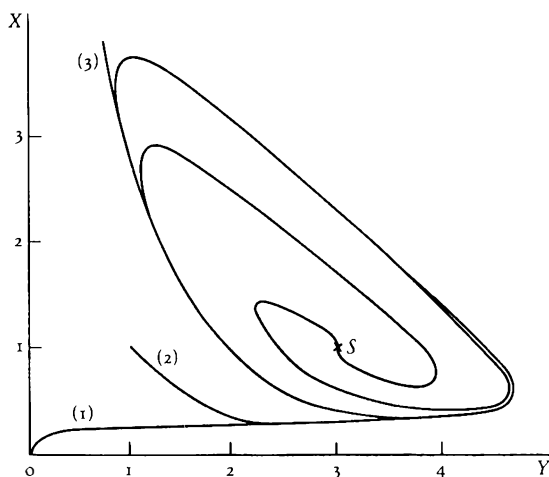
$$\begin{aligned} A &\rightarrow X \\ B + X &\rightarrow Y + D \\ 2X + Y &\rightarrow 3X \\ X &\rightarrow F \end{aligned} \tag{3.18}$$

Si assume che le concentrazioni di A e B siano controllate dall'esterno. Nella condizione limite di un sistema spazialmente uniforme (ben mescolato) le equazioni delle derivate divengono

$$\frac{dX}{dt} = A - BX + X^2Y - X$$
(3.19)
$$\frac{dY}{dt} = BX - X^2Y$$

Uguagliando a zero le derivate temporali si trova facilmente che il sistema possiede un unico punto fisso, $X_s = A$ e $Y_s = B/A$. Inoltre con metodi analitici standard, si mostra che questo punto è un fuoco stabile (cfr. figura 3.12*b*) se B è leggermente inferiore a $A^2 + 1$, e un fuoco stabile se B è leggermente superiore a $A^2 + 1$. In quest'ultimo caso esiste un ciclo limite con soluzione, illustrato nella figura 3.15. Le sue caratteristiche sono esattamente come quelle descritte con l'astratto modello matematico usato nel § 1. In altre parole, il valore critico

$$B_c = A^2 + 1 \qquad (3.20)$$

è un punto di biforcazione di Hopf. In effetti, l'analogia fra l'equazione 3.19 e il modello astratto matematico si spinge molto piú in là, ma intraprenderemo questa discussione piú avanti nel corso del paragrafo.

Figura 3.15
Ciclo limite calcolato numericamente per il modello a brussellatore (equazione 3.19). Valori dei parametri: $A = 1$; $B = 3$.

Un'altra classe di sistemi a due variabili che dà origine a cicli limite è costituita dai circuiti elettrici. Qui la non-linearità nasce dalla presenza di *elementi attivi* (valvole elettroniche, transistori, ecc.) che inducono caratteristiche corrente-tensione non-lineari. Un esempio classico è l'accoppiamento induttivo di un triodo e un circuito risonante, studiato inizialmente dallo scienziato olandese Balthazar Van der Pol. L'applicazione delle ben note leggi della conservazione della carica per i circuiti porta a una equazione chiusa per la tensione x sull'induttanza del circuito risonante, che, espressa in variabili adimensionali, è:

$$\frac{d^2x}{dt^2} - \varepsilon(1 - x^2)\frac{dx}{dt} + x = 0 \tag{3.21}$$

dove il parametro ε è collegato alle caratteristiche sia del triodo che del circuito risonante. L'analisi di questa equazione nel classico lavoro di Van der Pol mostrò l'esistenza di un'oscillazione prolungata di tipo a ciclo limite. In elettronica l'apparizione di tali oscillazioni è un ben noto fenomeno, anche se indesiderato, chiamato autoeccitazione.

Ma l'impatto delle biforcazioni descritto nei § 4 e 5 va al di là della loro applicabilità a sistemi a una o due variabili. Nella maggior parte dei problemi fisici si deve affrontare fin dall'inizio un grande numero di variabili (circa 10 nella reazione di Belousov-Žabotinski!), cosí se queste biforcazioni fossero limitate ai rari casi eccezionali discussi piú sopra, esse costituirebbero delle curiosità piuttosto che potenti mezzi di analisi dei sistemi naturali.

Fortunatamente non è cosí. Ci sono due meccanismi del tutto generali (ma non completamente irrelati), grazie ai quali la formulazione in termini di cuspidi, di punti limite, e delle biforcazioni di Hopf vista nei §§ 4 e 5 si applica a classi estremamente ampie di sistemi, indipendentemente dal numero delle variabili e dal grado di non-linearità coinvolti.

Iniziamo con il meccanismo concettualmente piú semplice, che dovrebbe evocare immediatamente concetti familiari a quei lettori che conoscono la chimica. In molti casi la dinamica di un sistema fisicochimico dà origine a una varietà di scale temporali fortemente diversificate dovuta a differenze nei valori dei parametri e/o dei valori delle variabili di stato. Per esempio, in chimica, una reazione catalitica sotto condizioni di laboratorio implica generalmente delle concentrazioni del catalizzatore che sono molto inferiori di quelle dei prodotti iniziali o finali, e velocità che sono molto piú alte di quelle intrinseche in assenza di un catalizzatore. Come risultato alcuni processi intermedi che coin-

volgono complessi catalitici procedono molto velocemente. Nella combustione le energie di attivazione di qualche reazione esotermica sono molto alte, cosí queste reazioni procedono, almeno negli stadi iniziali, molto piú lentamente del trasporto di energia e di momento. Esistono esempi simili in biologia, ottica, e in altri campi.

Intuitivamente, ci aspettiamo che durante un intervallo di tempo in cui le «variabili lente» cambiano faticosamente, le «variabili veloci» evolveranno rapidamente e raggiungeranno il livello previsto dalla soluzione dello stato stazionario delle loro equazioni di evoluzione, ma con una principale differenza: i valori delle variabili lente, che quelle veloci «vedranno» durante questo stadio non saranno i loro livelli di stato stazionario finale, ma, piuttosto, i valori istantanei previsti dalle equazioni del moto lento. Indicando con Y e X le variabili veloci e lente rispettivamente, ci aspettiamo dunque una relazione della forma

$$Y = h(X) \tag{3.22a}$$

che gioca essenzialmente il ruolo di un'«equazione di stato». Sostituendo questa espressione nelle equazioni per le variabili lente, in cui intervengono sia X che Y, si ottiene:

$$\frac{dX}{dt} = F(X, Y) = F[X, h(X)] = f(X) \tag{3.22b}$$

che costituisce un insieme chiuso di equazioni per X. Siamo dunque riusciti a ridurre considerevolmente il numero di variabili coinvolte eliminando quelle veloci. Tutto ciò è analogo alle idee che stanno dietro al «passo di velocità limitante» e all'«approssimazione di stato quasi stazionario» familiare in chimica e, piú specificatamente, nella cinetica degli enzimi. Se il numero delle rimanenti variabili lente è uno o due, cosa che spesso succede, si possono applicare i risultati dei §§ 4 e 5 a sistemi che sono molto piú generali degli astratti modelli matematici lí considerati. Un esempio esplicito sarà dato nel capitolo VI, § 3.

Nella nostra discussione abbiamo ripetutamente fatto appello all'intuizione, ma grazie a un fondamentale teorema dimostrato dal matematico russo Tichonov, in cui vengono date esplicitamente le particolari condizioni che le funzioni di produzione e i parametri debbono soddisfare, possiamo esprimere tutte le nostre principali conclusioni in termini rigorosi.

Discutiamo ora il secondo e piú sottile meccanismo che permette la

riduzione di un sistema iniziale a molte variabili a un sistema a poche variabili. Sia dato un insieme di equazioni di evoluzione della forma 3.1, in cui il numero n delle variabili coinvolte può essere molto alto. Con un metodo standard, noto come *analisi della stabilità lineare*, si possono determinare i valori dei parametri λ nei quali un certo stato di riferimento $\{X_{js}\}$ passa dalla stabilità asintotica all'instabilità. I lettori interessati in dettagli tecnici troveranno nell'*Appendice* A una presentazione compatta di questo metodo accompagnata da esempi di semplici modelli. Qui esporremo le idee principali.

Come discusso nel capitolo II, § 6, la stabilità è essenzialmente determinata dalla risposta del sistema alle perturbazioni o alle fluttuazioni che agiscono su uno stato di riferimento. È dunque naturale scrivere le leggi della dinamica, equazione 3.1, in una forma in cui compaiono esplicitamente le perturbazioni. Sostituendo

$$X_i(t) = X_{is} + x_i(t) \tag{3.23}$$

nella equazione 3.1 e ricordando che X_{is} è anche una soluzione di queste equazioni, arriviamo all'equazione

$$\frac{dx_i}{dt} = F_i(\{X_{is} + x_i\}, \lambda) - F_i(\{X_{is}\}, \lambda)$$

Queste equazioni sono omogenee nel senso che il membro a destra si annulla se tutti gli $x_i = 0$. Per avere una forma piú trasparente di questo sistema omogeneo, sviluppiamo $F_i(\{X_{is} + x_i\}, \lambda)$ attorno a $\{X_{is}\}$ e scrivendo esplicitamente la parte del risultato che è lineare in $\{x_j\}$, piú una correzione non-lineare di cui non serve per ora specificarne la struttura:

$$\frac{dx_i}{dt} = \sum_j L_{ij}(\lambda)x_j + h_i(\{x_j\}, \lambda) \qquad i = 1, \ldots, n \tag{3.24}$$

L_{ij} sono i coefficienti della parte lineare e h_i i contributi non-lineari. L'insieme degli L_{ij} definisce un *operatore* (una matrice $n \times n$ nel nostro caso), che dipende dallo stato di riferimento X_s e dai parametri λ.

Ora, un risultato fondamentale della teoria, discusso in maggior dettaglio nell'*Appendice* A, stabilisce che la stabilità asintotica o l'instabilità dello stato di riferimento $X = X_s$ (oppure $x = 0$) del sistema 3.24 sono identiche a quelle della sua parte linearizzata:

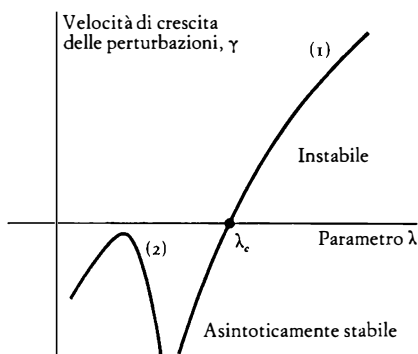
$$\frac{dx_i}{dt} = \sum_j L_{ij}(\lambda)x_j \qquad i = 1, \ldots, n \tag{3.25}$$

La stabilità si riduce in questo modo a un problema lineare risolvibile con i metodi dell'analisi elementare. È solo nel caso alla frontiera in cui $X = X_s$ (o $x = 0$) è stabile secondo Ljapunov ma non asintoticamente stabile, che la linearizzazione può non essere inammissibile. La stabilità si riduce cosí a un problema lineare, che è risolubile con metodi di calcolo elementare.

La figura 3.16 riassume il tipico risultato di una analisi di stabilità condotta secondo questa procedura. Ciò che si ottiene è il calcolo della velocità di crescita della perturbazione γ in funzione di un parametro (o piú) di controllo. Se $\gamma < 0$ (come è nella figura 3.16, ramo 1, se $\lambda < \lambda_c$) lo stato di riferimento è asintoticamente stabile, e se $\gamma > 0$ ($\lambda > \lambda_c$ nel ramo 1 della figura) esso è instabile. Per $\lambda = \lambda_c$ si ha uno stato di ***stabilità marginale***, la frontiera fra la stabilità asintotica e l'instabilità.

In generale un sistema in piú variabili dà luogo a un intero spettro di γ, proprio come un cristallo ha una moltitudine di modi di vibrazione. Si hanno dunque molte curve γ in funzione di λ nella figura 3.16. Supponiamo dapprima che fra tutte queste curve solo una (curva 1 in figura 3.16) attraversi l'asse delle λ, mentre tutte le altre sono sotto di essa. Con condizioni ben definite e deboli (mild) condizioni, discusse nell'*Appendice* A, si può mostrare che per $\lambda = \lambda_c$ *si verifica una biforcazione di nuovi rami di soluzione*. Si possono distinguere due casi:

Figura 3.16

Velocità di crescita della perturbazione γ in funzione del parametro di controllo λ, dedotto dall'analisi della stabilità lineare (cfr. equazione 3.25). Curva (1): lo stato di riferimento è asintoticamente stabile per $\lambda < \lambda_c$ e diventa instabile per $\lambda > \lambda_c$, dove λ_c è il valore critico della stabilità marginale. Curva (2): lo stato di riferimento rimane asintoticamente stabile per tutti i valori di λ.

1) Se per $\lambda = \lambda_c$ le perturbazioni non sono oscillatorie (questo è ciò che accade per esempio nella figura 3.12*a*), i rami di biforcazione corrisponderanno a soluzioni di stato stazionario.

2) Se per $\lambda = \lambda_c$ le perturbazioni sono oscillatorie (vedere per esempio la figura 3.12*b*), i rami della biforcazione corrisponderanno a soluzioni periodiche sotto forma di cicli limite.

In entrambi i casi, si può definire un opportuno insieme di quantità che obbediscono a un insieme chiuso di equazioni se i parametri rimangono vicini al loro valore critico λ_c. Nel caso 1 c'è solo *una* tale quantità, che misura l'ampiezza dei rami di biforcazione. Nel caso 2, d'altra parte, ci sono *due* di queste quantità che caratterizzano sia l'ampiezza che la fase dell'oscillazione. Effettivamente dunque la dinamica originale è sdoppiata in una singola equazione o un paio di equazioni che dànno informazioni sulla biforcazione e $n-1$ o $n-2$ equazioni che risultano essere «irrilevanti» per quanto riguarda la biforcazione. Chiamiamo *parametri d'ordine* le quantità che soddisfano le equazioni di biforcazione.

Un punto ancora piú sorprendente è che le equazioni di biforcazione risultano avere esattamente la stessa forma dei modelli matematici astratti, che dànno luogo alla biforcazione a forcone e di Hopf dei §§ 4 e 5. In altre parole, ogni dinamica che soddisfa l'una o l'altra delle precedenti condizioni 1 o 2 di cui sopra si può scrivere in una *forma normale* universale vicino al punto di biforcazione. Questo potente risultato era in effetti noto a Poincaré, ma fu il matematico americano George Birkhoff che lo sviluppò in profondità nello studio delle biforcazioni nei sistemi conservativi. Per i sistemi dissipativi l'idea della riduzione a pochi parametri è associata frequentemente al nome del fisico sovietico Lev Landau, in connessione con la sua teoria delle transizioni di fase. Nei testi di matematica è frequentemente citata come procedura di Ljapunov-Schmidt e, piú recentemente, come *teoria della varietà centrale*. Il lettore interessato troverà una derivazione compatta delle equazioni di biforcazione nell'*Appendice* B.

Si possono anche immaginare situazioni piú intricate nelle quali piú rami attraversano l'asse delle λ nella figura 3.16. Questo porta all'interazione tra soluzioni biforcanti che genera fenomeni di biforcazione secondari o perfino terziari. I risultati precedenti rimangono validi, nel senso che si può garantire che la parte di dinamica che dà l'informazione sui rami di biforcazione si svolge in uno spazio delle fasi di dimensio-

ne piú bassa. La costruzione esplicita della forma normale è comunque molto piú involuta e la sua universalità non può piú essere garantita.

Riassumendo, dunque, si può affermare che le biforcazioni statiche e dinamiche passate in rassegna nei §§ 4 e 5 sono autentici *archetipi di complessità*, nel senso che sono presenti in una vasta classe di sistemi e sotto ben differenti condizioni. Nel prossimo paragrafo vedremo che, nonostante la loro generalità, esse costituiscono solo una parte dello stupefacente repertorio dei comportamenti dei sistemi dinamici.

7. *Lo spazio delle fasi rivisitato: varietà topologiche e frattali.*

Abbiamo osservato che in tutti i sistemi dove la dinamica è riducibile in uno spazio delle fasi a una o due dimensioni, il comportamento è fortemente limitato da vincoli topologici imposti a ogni moto in poche dimensioni. In particolare risulta che i soli possibili attrattori sono i punti fissi ed i cicli limite.

Ci interesseremo ora delle proprietà dei sistemi dinamici che si evolvono in spazi delle fasi la cui dimensione è piú grande di due. Alla fine del § 1, abbiamo sottolineato che è probabile che la libertà data da dimensioni piú grandi sia all'origine di comportamenti completamente nuovi. Indipendentemente da questi nuovi aspetti, pensiamo di riottenere tutte le caratteristiche dei sistemi a una o due dimensioni, e, in particolare, gli attrattori punto fisso e ciclo limite. Naturalmente, la struttura topologica delle traiettorie attorno a essi sarà piú complicata, ma se si attende sufficientemente a lungo, alla fine si arriverà a uno stesso comportamento. Lasciamo dunque questo problema da parte e volgiamo la nostra attenzione alle nuove possibilità.

Per ottenere la piú chiara descrizione possibile, intraprenderemo la nostra discussione per sistemi che evolvono in uno spazio delle fasi tridimensionale. Nella nostra analisi esplorativa, vogliamo naturalmente beneficiare dell'esperienza che abbiamo già acquisito nel trattare i problemi bidimensionali. Proviamo a individuare una buona strategia. Una possibilità ci è suggerita direttamente dalla *Repubblica* di Platone con il celebre mito della caverna: guardare l'«ombra» di un intero oggetto, e cercare di ricostruirlo dal suo aspetto bidimensionale. Platone stesso ammette che questa è un'impresa molto difficile, e solo la filosofia può portare la mente umana vicino a questo scopo. In termini piú moderni sappiamo che una particolare proiezione non è certamente in corrispondenza biunivoca con l'oggetto. Sono necessarie molte proiezioni su su-

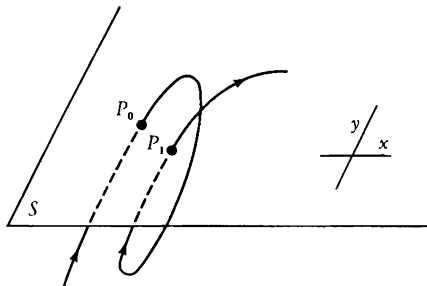
perfici diversamente orientate per descriverlo. Ma questo ci porta automaticamente allo spazio multidimensionale originale, che è precisamente ciò che vogliamo evitare.

Fortunatamente vi è una seconda possibilità suggerita da un altro grande genio, Henri Poincaré. Consideriamo un piano S che taglia trasversalmente le traiettorie dello spazio delle fasi (figura 3.17) e studiamo il moto dei punti di intersezione P_n delle traiettorie. A seconda dell'oggetto che attrarrà i punti P_n per lungo tempo, saremo in grado di inferire l'attrattore racchiuso nello spazio delle fasi tridimensionale, perché avremo di fatto a disposizione una sezione di questo attrattore. Per questa ragione S è chiamata superficie di sezione di Poincaré, e la dinamica dei P_n la *mappa di Poincaré*. Osserviamo che questa dinamica è infatti una *iterazione* in cui il tempo interviene in modo discreto, poiché gli intervalli fra le intersezioni sono finiti. Introducendo un appropriato sistema di coordinate nel piano S, possiamo scrivere questa iterazione nella forma

$$\begin{aligned} x_{n+1} &= f(x_n, y_n) \\ y_{n+1} &= g(x_n, y_n) \end{aligned} \tag{3.26}$$

Ora che le dinamiche sono state ridotte a una dinamica bidimensionale possiamo applicare i risultati precedenti e rendere plausibili le congetture su di essa. Ci proponiamo di fare una preliminare classificazione delle varie possibilità.

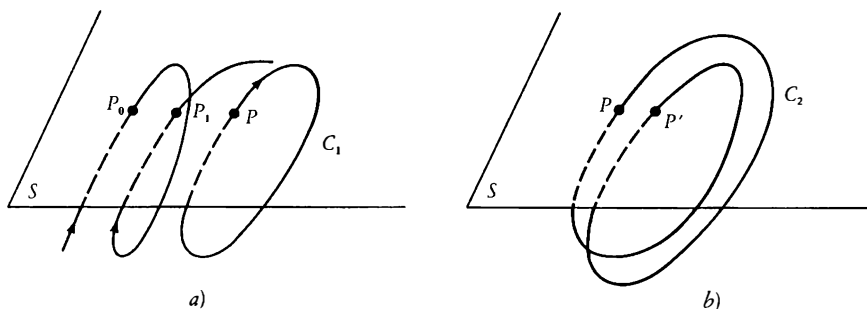
Figura 3.17

Superficie di sezione di Poincaré. Sono studiate le intersezioni delle traiettorie di un sistema dinamico che evolve continuamente nel tempo, con una superficie S trasversale alla traiettoria. La dinamica temporale discreta che descrive la posizione dei successivi punti di intersezione P_0, P_1, ... su S dà un'ampia informazione sul sottostante flusso continuo.

Attrattori periodici: cicli di ordine k.

La figura 3.18*a* suggerisce la possibilità piú ovvia: i punti di intersezione P_0, P_1, ..., convergono a un punto fisso P. Per orbite limitate ciò ovviamente significa che l'attrattore è un ciclo limite, C_1. Nella relazione di ricorrenza 3.26 l'esistenza di P mostra la possibilità di soddisfare contemporaneamente le equazioni $x_{n+1} = x_n$, $y_{n+1} = x_n$.

Altre possibilità sono tuttavia permesse. Invece del «ciclo di ordine uno» definito ora, si può richiedere che gli x_{n+1}, ecc. rimangano diversi da x_n (analogamente per y), finché si raggiunge una iterazione k per la quale $x_{n+k} = x_n$, $y_{n+k} = y_n$. Questo viene chiamato «ciclo di ordine k». La figura 3.18*b* descrive come dovrebbe apparire un ciclo di ordine due. Nella superficie di Poincaré l'attrattore è un insieme di due punti P e P' successivamente visitati dal sistema, mentre l'oggetto corrispondente nello spazio tridimensionale è una curva C_2 che si ritorce prima di chiudersi su se stessa. Questo sarebbe impossibile in uno spazio delle fasi bidimensionale poiché implicherebbe l'autointersezione delle traiettorie. Chiaramente questo è il modello naturale per descrivere i fenomeni periodici il cui periodo diventa improvvisamente piú lungo al variare di uno dei parametri. Tali transizioni sono state abbondantemente osservate sperimentalmente in fluidomeccanica, chimica od ottica, e saranno ulteriormente discusse nel § 10.

Figura 3.18

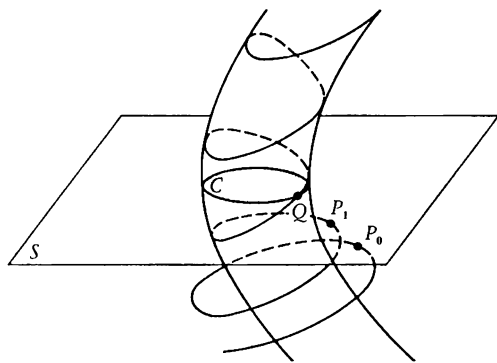
Convergenza delle traiettorie di un sistema continuo dinamico verso un ciclo limite C_1 e un ciclo limite C_2 il cui periodo è doppio del periodo di C_1. Nella superficie di sezione di Poincaré ciò risulta evidente dalla convergenza della sequenza P_0, P_1, ... rispettivamente verso un punto fisso stabile P e verso una coppia di punti P, P'.

Si noti che, come nel § 5, i punti fissi della mappa di Poincaré possono anche essere classificati come selle, nodi, ecc.

Attrattori quasi periodici: tori invarianti.

Un'altra possibilità topologicamente ammissibile è suggerita dalla figura 3.19: i punti di intersezione P_0, P_1, ..., convergono a una curva chiusa C nella superficie di Poincaré S. Il corrispondente oggetto attraente nello spazio tridimensionale è dunque una superficie bidimensionale, e, se richiediamo su basi fisiche che le traiettorie non vadano all'infinito, la superficie deve essere limitata. Inoltre la sua struttura deve permettere alle traiettorie di evolvere senza intersezione, come richiesto dal teorema di unicità descritto nel § 1. Questo esclude la sfera e ci lascia con il *toro*, già incontrato nel § 3. Come spiegato in quest'ultimo paragrafo il moto su una superficie toroidale è generalmente quasi periodico, con due frequenze indipendenti e incommensurabili.

Il nuovo elemento qui è che il toro diventa attrattore, contrariamente al caso dei sistemi conservativi discussi nel § 3 in cui c'era un continuum di tori che circondavano un punto ellittico di equilibrio.

Figura 3.19

Attrattore quasi periodico sotto forma di un toro bidimensionale T_2 immerso in uno spazio delle fasi tridimensionale. Durante lo stadio transiente di evoluzione la traiettoria interseca la superficie di sezione di Poincaré di S sui punti P_0, P_1, ..., prima di convergere alla superficie del toro. In seguito continua ad avvolgersi indefinitamente su tale superficie. A ogni passaggio attraverso S la traiettoria dà luogo a un punto di intersezione come Q. L'insieme di tutte le tracce di successivi passaggi attraverso S costituisce la curva C, che è l'attrattore della superficie di Poincaré verso cui converge la sequenza dei punti transienti P_0, P_1, ...

In altre parole un toro attrattore invariante è il modello naturale dei fenomeni quasi periodici e riproducibili osservati in natura.

Attrattori non periodici: frattali.

Possono esistere altri attrattori in uno spazio delle fasi tridimensionale oltre ai punti fissi, le curve chiuse o le superfici limitate bidimensionali? Notiamo dapprima che un attrattore deve necessariamente avere un numero di dimensioni piú basso di quello dello spazio delle fasi, poiché esso rappresenta un insieme di stati che rimane invariante per la dinamica. La domanda ora sollevata equivale dunque a chiedere se esistono oggetti il cui numero di dimensioni è fra quello di un punto e una linea, fra quello di una linea e una superficie, o persino fra quello di una superficie e un volume. Tali oggetti, se esistono, non saranno né punti, né curve, né superfici, né, piú in generale, «varietà topologiche». Il matematico francese Benoît Mandelbrot ha dato loro il nome di *frattali*.

La teoria dei numeri, generalmente guardata come una delle parti piú profonde della matematica, ci permette di prevedere certe possibilità poiché la questione sollevata si può ridurre alla seguente: esistono degli insiemi che sono intermedi fra un insieme numerabile come, per esempio, i numeri naturali, e un insieme continuo, come, per esempio, i punti di una linea? La risposta a questa domanda è affermativa. È strettamente collegata alla nozione di infinito che venne fornita dal matematico tedesco Georg Cantor. Non possiamo entrare qui nei dettagli del ragionamento matematico, e cosí illustreremo solo il risultato con un particolare esempio noto come *insieme di Cantor* (figura 3.20).

Si consideri l'intervallo chiuso [0, 1]. Lo si divida in tre parti uguali togliendo poi la parte centrale senza gli estremi. In ognuno dei rimanenti intervalli, che pure abbiamo diviso in tre parti uguali, si tolga nuovamente l'intervallo aperto centrale, e cosí di seguito. Per costruzione l'insieme dei punti ottenuto è un insieme infinito, numerabile di

Figura 3.20

Costruzione dell'insieme di Cantor. Le sezioni in grigio sono le parti centrali eliminate in successive divisioni in tre dei segmenti della linea.

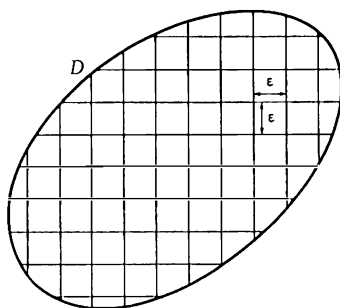
punti non connessi. Non ha una lunghezza intrinseca (in termini piú tecnici ha misura nulla), poiché la lunghezza del suo complementare è uguale a uno: $\frac{1}{3} + \frac{2}{9} + \frac{4}{27} \cdots = 1$. Sorprendentemente la sua dimensione non è zero! Vediamo come questo può accadere.

Come esprimiamo in termini astratti che il pezzo di carta su cui scriviamo è bidimensionale, o che lo spazio nel quale ci muoviamo è tridimensionale? Qualunque sia l'insieme di interesse, lo si chiami D e si provi a riempirlo di segmenti, quadrati, o, piú in generale, ipercubi il cui lato sia uguale a ε. Si osservi quindi che tutte le dimensioni d delle ben note superfici «topologiche» si deducono dalla seguente relazione:

$$d = \lim_{\varepsilon \to 0} \frac{\ln N_\varepsilon}{\ln 1/\varepsilon} \tag{3.27}$$

Il concetto è illustrato dalla figura 3.21. Scegliamo una partizione in celle di lunghezza caratteristica uguale a ε, contiamo il numero minimo di celle N, necessario a coprire D, dividiamo il logaritmo di N per il logaritmo di $1/\varepsilon$, ne facciamo il limite per ε molto piccolo, e otteniamo δ la dimensione di D.

Applichiamo questa procedura al piano, specificatamente al quadrato di lato 1. Sono necessari $(1/\varepsilon)^2$ quadratini di lato ε per coprire questo spazio, e quindi abbiamo $\ln (1/\varepsilon)^2/\ln (1/\varepsilon) = 2\ln (1/\varepsilon)/\ln (1/\varepsilon) = 2$, cioè $d = 2$, come ci si aspettava. Ma cosa si ottiene se estendiamo questo concetto all'insieme di Cantor?

Figura 3.21
Illustrazione del concetto di dimensionalità, d, di un insieme D. Il numero minimo di ipercubi di grandezza ε necessari per coprire completamente l'insieme viene conteggiato per diversi valori della grandezza ε, e viene valutato il limite dell'equazione 3.27.

Nella prima partizione della linea due segmenti di lunghezza 1/3 sono sufficienti per descrivere tutti i sottoinsiemi presenti. Nella seconda partizione sono necessari quattro segmenti di lunghezza 1/9, e piú generalmente, nella n-esima partizione 2^n segmenti di lunghezza $1/3^n$. Si ha dunque $\ln N_\varepsilon / \ln (1/\varepsilon)^2 = \ln 2^n / \ln 3^n = (n \ln 2)/(n \ln 3)$, e quindi

$$d_{\text{Cantor}} = \frac{\ln 2}{\ln 3} \simeq 0.63 \tag{3.28}$$

L'insieme di Cantor è dunque, sotto questo aspetto, «intermedio» tra un punto ($d = 0$) ed una linea ($d = 1$): è un *oggetto frattale.*

Stabilita quindi l'esistenza di questi insiemi, possiamo ora immaginare attrattori nello spazio delle fasi tridimensionale che si presentano nella superficie di sezione di Poincaré come una infinità non numerabile di segmenti non connessi, alternantesi tra loro in un modo non riproducibile mentre il tempo segue il suo corso. L'oggetto corrispondente sarà costituito quindi da un numero infinito di fogli tale che una sezione trasversale ai fogli sia un insieme di Cantor. È naturale aspettarsi che tali oggetti, frequentemente denominati *strani attrattori*, costituiranno i modelli naturali per descrivere l'inizio del comportamento caotico, di tipo turbolento.

Abbiamo centrato la nostra discussione sugli attrattori che possono esistere solo nei sistemi dissipativi, ma chiaramente tutti gli argomenti citati possono essere usati altrettanto bene per descrivere gli insiemi invarianti di un sistema conservativo. Il nostro compito nei tre successivi paragrafi sarà quello di convincerci che le possibilità sopra menzionate, le quali a questo livello sono semplici speculazioni motivate da considerazioni topologiche, corrispondano di fatto alla realtà fisica.

8. *Sistemi conservativi non integrabili: la nuova dinamica.*

I sistemi conservativi integrabili considerati nel § 3 costituiscono una classe piuttosto eccezionale. Pensiamo per un momento al movimento dei pianeti attorno al Sole. La massa dei pianeti è circa mille volte inferiore della massa del Sole, cosí in prima approssimazione possiamo trascurare la loro interazione e considerare solamente la loro attrazione da parte del Sole. Otteniamo in questo modo il *problema dei due corpi*, un classico esempio di un sistema integrabile (ed esattamente risolubile) che dà luogo a un moto periodico per ogni singolo pianeta, e a un moto quasi periodico per l'intero sistema planetario.

L'interazione fra pianeti, per quanto piccoli possano essere, è sempre presente, e tende a perturbare le traiettorie kepleriane. Il problema è se questa perturbazione induce solo deboli aggiustamenti quantitativi oppure se, al contrario, modifica gli aspetti qualitativi del moto. Questo è il problema fondamentale della meccanica celeste. Si trovano problemi molto simili a questo in una grande varietà di altri sistemi meccanici.

Chiaramente, essendo interessati a scale temporali dell'ordine dell'inverso della forza della perturbazione (circa mille anni per il nostro sistema planetario), possiamo esaminare gli effetti indotti da quest'ultima in modo diretto. È sufficiente cercare soluzioni approssimate per le equazioni complete aggiungendo una piccola correzione, proporzionale alla forza della perturbazione, al termine di ordine zero che rappresenta il sistema integrabile «di riferimento». Questo si può fare con i metodi usuali. Ma se cerchiamo soluzioni valide per tempi lunghi e, in particolare, se vogliamo capire il comportamento qualitativo delle soluzioni esatte, ci imbattiamo immediatamente in grandissime difficoltà. Nell'esempio del sistema planetario queste sorgerebbero in tempi dell'ordine del miliardo di anni, e la domanda finale può essere se i pianeti fuggiranno dal sistema, se cadranno sul Sole o se si scontreranno gli uni con gli altri. Domande meno drammatiche, ma tecnicamente molto simili, sorgono nel moto di particelle cariche veloci in intensi campi magnetici come quelli prevalenti negli anelli di confinamento dei giganteschi acceleratori di particelle, o nella stabilità orbitale di una navicella spaziale.

La principale manifestazione delle difficoltà sopra citate è la presenza di *divergenze*. In altre parole, valutando con i tradizionali metodi di perturbazione la correzione al moto di un sistema integrabile, si trova che essa diventa molto grande o persino sotto certe condizioni tende all'infinito. Tra queste manifestazioni la piú diffusa è la *risonanza*, ovvero, il caso in cui le frequenze del sistema di riferimento integrabile sono commensurabili nel senso di $k_1\omega_1 + k_2\omega_2 + \cdots = 0$ (cfr. § 3). Questo segnala il fallimento del metodo usato e suggerisce fortemente che ci dovrebbe essere un cambiamento qualitativo nel comportamento del sistema sotto queste condizioni.

Cerchiamo di capire, su basi intuitive, i due aspetti della difficoltà. Primo: perché le risonanze sono abbondanti anziché eccezionali; secondo, perché esse sono «pericolose»?

La teoria dei numeri ci dice che gli irrazionali costituiscono la grande maggioranza dei numeri reali, poiché, da un lato, essi costituiscono

un insieme non numerabile (contrariamente ai razionali) ed essi supportano l'intera «lunghezza» o «misura» di ogni segmento sulla retta reale. Cosí (cfr. equazione 3.8) nella maggior parte dei punti dello spazio delle fasi le variabili di azione $\{I_i\}$ saranno tali che le frequenze $\{\omega_i\}$ realizzate dal sistema saranno collegate in modo irrazionale e dunque saranno incommensurabili. La difficoltà sopra citata semplicemente non sorgerà, e il moto avrà luogo di regola su *tori non risonanti* (cfr. figura 3.7*b*) generati da un tale insieme di frequenze irrazionalmente collegate. Ma le cose non sono cosí semplici. Anche se i numeri razionali sono «eccezionali» in un certo senso, essi rimangono sufficientemente «abbondanti» (nella terminologia matematica si dice che i razionali sono *densi* nell'insieme dei numeri reali), poiché in ogni prescritto intorno di un numero reale esistono sempre dei numeri razionali. In particolare, ogni *volume finito* dello spazio delle fasi, per quanto piccolo, conterrà un numero infinito di punti rappresentativi in cui è soddisfatta la condizione di risonanza. Cosí ci si aspetta che le risonanze giochino un ruolo importante nell'evoluzione del sottostante insieme di Gibbs (cfr. § 2) e, attraverso di esso, nella descrizione statistica del sistema.

Vediamo ora perché è probabile che le risonanze siano «pericolose». Consideriamo di nuovo la relazione 3.10 in connessione con la seconda equazione 3.8. Ciò che esse implicano realmente è che esiste una particolare combinazione di variabili d'angolo, $\varphi = k_1\varphi_1 + \cdots + k_n\varphi_n$ che rimane invariante durante il moto e può anche venire eliminata dalla descrizione. In altre parole, la nostra scelta iniziale di n variabili indipendenti ci induceva in errore poiché solo $n - 1$ di loro sono veramente indipendenti. In termini piú tecnici il nostro toro a n dimensioni si riduce a uno di $n - 1$ dimensioni. Chiamiamo tali superfici invarianti a basse dimensioni *tori risonanti*. Per due gradi di libertà questo significa che la superficie invariante è coperta da famiglie di orbite chiuse (unidimensionali) che descrivono un moto periodico anziché quasi periodico. Una di queste orbite genera su una superficie di sezione di Poincaré S – ad esempio la curva C nella figura 3.22 – un numero di punti fissi (P_1 e P_2 nella figura 3.22) uguale al numero di giri che compie prima di chiudersi, in una direzione trasversale alla superficie della sezione. Chiamiamo quest'ultimo numero il *numero di avvolgimento* (*winding number*) dell'orbita.

Ogni altro membro della famiglia genererà naturalmente un altro insieme di un ugual numero di punti fissi che giacciono su S. Cosí, l'intera curva C sarà fatta di punti fissi. È difficile evitare la sensazione che questa sia una situazione eccezionale, paragonabile, per esempio, all'a-

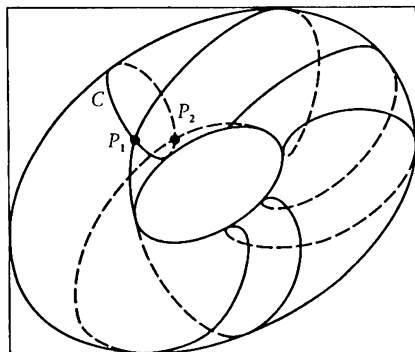
vere tre punti scelti a caso che giacciono su una linea retta. C'è dunque da aspettarsi che le perturbazioni piú tipiche (non integrabili) a cui possono essere soggetti i sistemi integrabili distruggeranno questa configurazione delle traiettorie nello spazio delle fasi. Ciò dovrebbe mostrarsi attraverso il fatto che la correzione al moto imperturbato sarà importante tanto quanto l'ampiezza di questo stesso moto.

La ricerca di una soluzione adatta per trattare le difficoltà sollevate dai sistemi non integrabili ha portato a una grossa rivoluzione scientifica, il cui risultato è stato il sorgere di una nuova meccanica a partire dagli anni '50. Molti degli ingredienti di questo sviluppo erano noti a Poincaré e Birkoff, ma il matematico russo Andreij Kolmogorov è considerato come il principale artefice di questa conquista. Successivamente sono stati portati importanti contributi da Vladimir Arnold e Jürgen Moser fra gli altri. Riassumiamo brevemente i punti principali.

Perturbazione di moti quasi periodici.

Le argomentazioni utilizzate piú sopra in questo paragrafo per suggerire che i tori risonanti debbono essere sensibili alle piccole perturbazioni non integrabili, ci inducono a pensare che nel caso non risonante

Figura 3.22

I punti fissi P_1 e P_2, generati sull'intersezione C tra la superficie di sezione di Poincaré e un toro risonante allorché una particolare traiettoria periodica si avvolge su di esso. Ogni altro membro del continuum delle traiettorie periodiche genererà un ugual numero di punti fissi sulla superficie di Poincaré: l'intera curva C sarà quindi fatta di punti fissi. Ci si può aspettare che questa situazione patologica sia distrutta da una tipica (non integrabile) perturbazione.

il moto debba rimanere «robusto». In altre parole, tipicamente, il comportamento quasi periodico non viene fortemente alterato da una debole perturbazione. Questo importante risultato è stato rigorosamente dimostrato da Kolmogorov, che, inoltre, è stato in grado di caratterizzare in un modo sottile i moti quasi periodici che sussistono.

Specificatamente, sembra che affinché sussista un moto quasi periodico le frequenze debbano essere «sufficientemente» irrazionali nel senso che segue (qui esemplificato per due gradi di libertà): il rapporto ω_1/ω_2 deve essere sufficientemente diverso in valore assoluto da un valore razionale,

$$\left| \frac{\omega_1}{\omega_2} - \frac{r}{s} \right| > \frac{c}{s^\nu} \qquad \text{per tutti gli interi } r \text{ e } s \tag{3.29}$$

dove c è un numero piccolo che si annulla in assenza della perturbazione, e ν è sufficientemente grande (diciamo $\nu = 2.5$). Si può mostrare che i moti quasi periodici permessi dall'equazione 3.29 costituiscono la maggioranza dei moti realizzati nello spazio delle fasi.

Concludendo, ci aspettiamo che in vaste regioni dello spazio delle fasi un sistema non integrabile vicino a uno integrabile dovrebbe mostrare un comportamento simile alla figura 3.8, ovvero, quello dei moti stabili che circondano orbite periodiche.

Perturbazione di moti periodici.

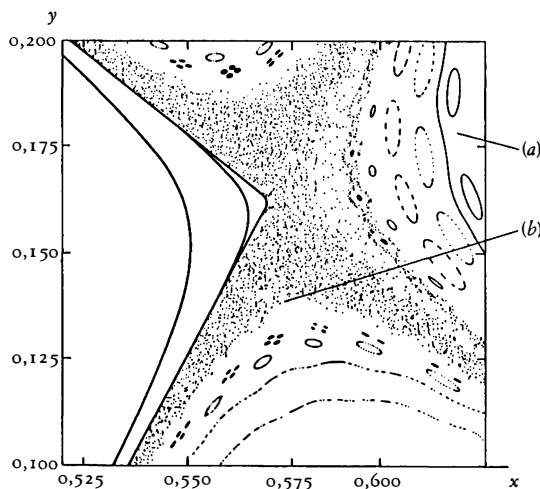
È conveniente affrontare questo problema con il metodo della superficie di sezione di Poincaré S (figura 3.22) e indagare dapprima sulla sorte del continuum dei punti fissi di S sotto l'effetto della perturbazione. Sia s il numero di avvolgimento. Si può dimostrare (vedere l'*Appendice* C per una spiegazione succinta) che la perturbazione dà origine a $2ks$ punti fissi (k è un certo intero), di cui ks sono ellittici e ks sono iperbolici (selle). I punti fissi ellittici corrispondono a orbite periodiche stabili, e ognuna di esse è circondata da tori invarianti (non risonanti) che appaiono nella mappa di Poincaré come curve chiuse. La corrispondente parte dello spazio delle fasi è frequentemente chiamata «isola» (figura 3.23, regione a).

La situazione è completamente differente per i punti fissi iperbolici. A causa dell'instabilità delle corrispondenti orbite periodiche, si sviluppa una struttura estremamente intricata nelle loro vicinanze, che è essenzialmente dovuta alla peculiare configurazione assunta dalle loro se-

paratrici (figura 3.23, regione *b*). In particolare, nella mappa di Poincaré queste separatrici si piegano e si intersecano in un numero infinito di punti noti come punti omoclinici (cfr. *Appendice* C). In questa regione compare un regime di *moto altamente irregolare*: le traiettorie sono molto sensibili a deboli cambiamenti delle condizioni iniziali e vagano in modo irregolare lungo ampie regioni dello spazio delle fasi. Nella mappa di Poincaré questo si manifesta attraverso la comparsa di una nuvola di punti che sembrano essere disposti a caso, anche se le traiettorie sono ancora, naturalmente, assolutamente deterministiche. Questo comportamento caotico suggerisce che ci deve essere una profonda relazione fra sistemi conservativi non integrabili e il comportamento stocastico che generalmente viene associato all'irreversibilità e ai sistemi dissipativi. Questo punto importantissimo verrà ripreso in dettaglio nel capitolo V.

Si noti che la struttura globale si ripete a tutti i livelli attorno ai punti fissi ellittici generati sulla superficie di Poincaré dalle varie orbite risonanti. Questa notevole autosomiglianza mostra l'enorme complessità raggiunta nei problemi apparentemente ordinari come due gradi di li-

Figura 3.23

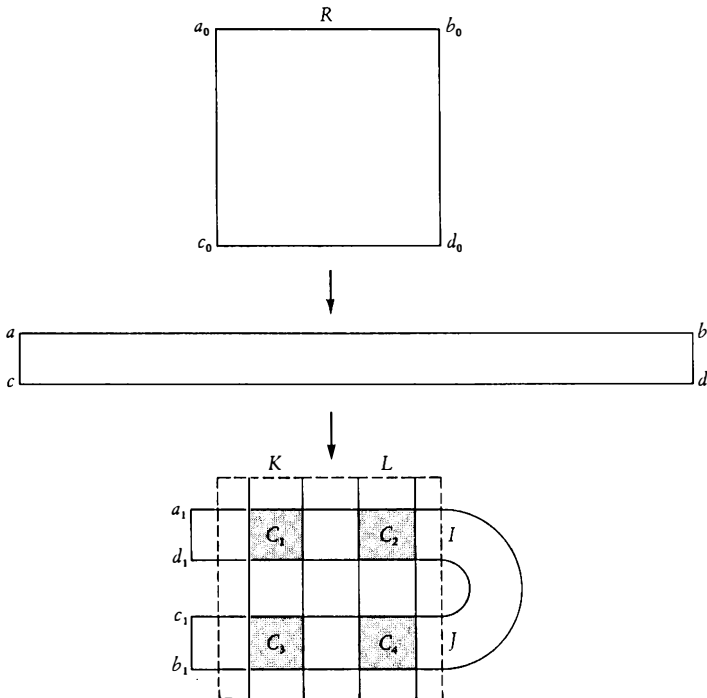
Traiettorie regolari e stocastiche per un insieme hamiltoniano non integrabile visualizzato su una superficie di Poincaré. Regione *a*): moto stabile vicino ad un punto fisso ellittico; regione *b*): moto instabile vicino ad un punto fisso iperbolico. (Da M. Hénon, *Q. Appl. Math.*, 27, 291, 1969).

bertà debolmente interagenti. Una gran mole di lavoro numerico con lo scopo di chiarire questo comportamento è stata portata avanti con semplici modelli. Per piccoli valori della perturbazione i calcoli corroborano la descrizione generale data nella figura 3.23. Ma non appena l'energia del sistema aumenta, osserviamo che le isole tendono a ritirarsi, e le traiettorie caotiche invadono progressivamente porzioni sempre piú vaste dello spazio delle fasi. È difficile trarre delle conclusioni qualitative da tali osservazioni, poiché questo regime è difficilmente trattabile con calcoli analitici o anche con generali considerazioni topologiche. Tuttavia è possibile introdurre una misura che caratterizzi l'instabilità del moto e la dipendenza dalle condizioni iniziali. Allo scopo, vi sono gli *esponenti di Ljapunov*, che descrivono la velocità media della divergenza esponenziale di due traiettorie inizialmente vicine. Gli esponenti di Ljapunov costituiscono la generalizzazione dei concetti della teoria della stabilità lineare, introdotti nel § 6, ai sistemi caotici. Per maggiori dettagli, si veda l'*Appendice* A.

La struttura complicata di traiettorie regolari e irregolari discusse piú sopra non è un aspetto esclusivo dei sistemi hamiltoniani non integrabili. È il comportamento tipico di altre ampie classi di sistemi conservativi, fra i quali i cosiddetti sistemi di Kolmogorov (sistemi K) giocano un ruolo importante. Rimandiamo la discussione dei sistemi K al capitolo v, poiché essi costituiscono il nostro strumento principale per illustrare il passaggio dalla dinamica conservativa a quella dissipativa. Un approccio molto istruttivo allo studio dei sistemi dinamici instabili si attua con trasformazioni temporali discrete, a causa della loro intima relazione, per mezzo della superficie di sezione di Poincaré (cfr. equazione 3.26), con i sistemi continui nel tempo. Infatti, a parte qualche rara eccezione, le trasformazioni discrete sono gli unici sistemi dinamici noti per i quali si possa stabilire esplicitamente e rigorosamente l'apparire di moti irregolari che mostrano un comportamento caotico. Un famoso esempio inventato dal matematico americano Stephen Smale è la *trasformazione a ferro di cavallo*, che si può formulare sia come un astratto sistema dinamico conservativo, sia come un sistema dinamico dissipativo. Sebbene nessun sistema hamiltoniano noto si possa ridurre esplicitamente a questa trasformazione, essa verrà descritta piú sotto e successivamente la utilizzeremo come archetipo di questo nuovo comportamento, cosí come abbiamo usato i cicli e i tori come archetipi di modi piú semplici di comportamento.

9. *Un modello di moto instabile: il ferro di cavallo.*

Consideriamo una mappa, T, del piano in se stesso che agisce su un quadrato R nel modo mostrato in figura 3.24. Il quadrato è compresso in direzione verticale ed è espanso in direzione orizzontale, e viene successivamente piegato fino a formare un ferro di cavallo, in modo tale che la parte piegata e gli estremi della «barra» originale si proiettano al di fuori del quadrato. In questo modo R non si trasforma in se stesso: solo le strisce I e J del ferro di cavallo sono contenute in R. Le loro «pre-immagini» (cioè, quelle parti di R che le generano sotto l'azione

Figura 3.24

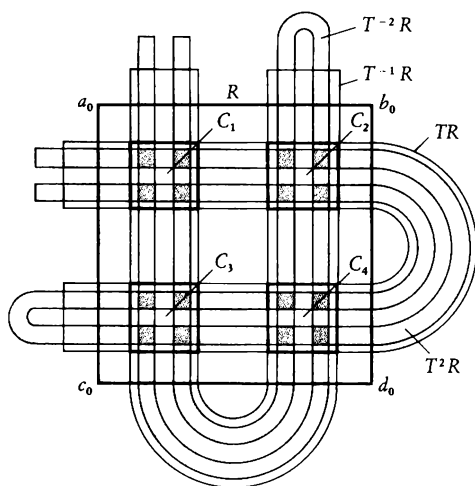
La mappa a ferro di cavallo. Il quadrato a_0 b_0 c_0 d_0 viene stirato di almeno un fattore 2 nella direzione orizzontale e compresso di un fattore piú grande di 2 nella direzione verticale, e viene quindi piegato fino a formare un ferro di cavallo. I quadrati da C_1 a C_4 rimangono invarianti rispetto alla trasformazione.

della mappa) sono in effetti due strisce verticali K e L del quadrato originario.

L'insieme invariante di punti Λ della mappa e il suo inverso T^{-1} sono ovviamente contenuti in entrambe le coppie di rettangoli I, J e K, L. Sono quindi un sottoinsieme dei quattro quadrati piú piccoli, da C_1 a C_4, della figura 3.24. Chiamiamo questi *insiemi iterativi.*

La figura 3.25 mostra il risultato di una seconda interazione, $T^2 = T \cdot T$, della mappa con R. In particolare essa mostra le strisce di $T^2 \cdot R$ che sono contenute nel quadrato originale e le loro pre-immagini, che sono parte dell'insieme generato dall'azione della mappa inversa T^{-2} su R. Otteniamo un insieme iterativo di sedici quadrati, quattro entro ogni quadrato da C_1 a C_4 della figura 3.24.

Continuando le iterazioni si vede che i punti che non lasceranno mai R sono situati nei 4^n quadrati le cui dimensioni decrescono esponenzialmente a zero al tendere di n all'infinito. L'insieme di punti ricorrenti Λ che ne risulta è composto da un'infinità di punti non numerabili, e la sua area totale è zero. Esso possiede dunque l'aspetto di un insieme di Cantor o, piú precisamente del prodotto di due insiemi di Cantor.

Una dettagliata analisi delle dinamiche nell'insieme Λ può essere

Figura 3.25
La mappa a ferro di cavallo iterata due volte, che illustra la formazione dell'insieme Λ di punti ricorrenti per successive frammentazioni dei quadrati da C_1 a C_4.

condotta tramite la tecnica delle *dinamiche simboliche*, argomento che sarà toccato più avanti. Si può mostrare in questo modo che qualunque intorno di ogni punto di Λ contiene orbite periodiche nel senso descritto dal § 7. Inoltre esistono orbite periodiche che si avvicinano arbitrariamente a ogni punto di Λ. Ogni punto di Λ è per costruzione un punto iperbolico, poiché è l'intersezione di una direzione che si contrae (la direzione verticale) e di una che si espande (la direzione orizzontale). Pertanto due orbite che partono da due punti vicini di Λ divergeranno esponenzialmente nel corso delle iterazioni successive. Questo implica una estrema dipendenza dalle condizioni iniziali, una delle principali caratteristiche delle traiettorie irregolari incontrate nei precedenti paragrafi in connessione con i sistemi hamiltoniani non integrabili, e degli strani attrattori considerati nel prossimo paragrafo.

Di fatto la precedente analogia va ben oltre: si può mostrare che, quando la separatrice dei punti instabili della mappa di Poincaré si torce e si interseca come nella figura 3.23, il moto originale genera localmente sulla superficie di sezione di Poincaré una dinamica di forma simile a un ferro di cavallo. Questo è il motivo dell'uso di questo modello come archetipo dei moti irregolari e, in concomitanza, del comportamento caotico.

10. *Sistemi dissipativi in spazi delle fasi multidimensionali. Caos e strani attrattori.*

Consideriamo ora le implicazioni delle diverse possibilità nel comportamento dei sistemi dissipativi prese in esame nei §§ 7, 8 e 9. Abbiamo visto che nei sistemi conservativi possono coesistere nello spazio delle fasi traiettorie regolari e caotiche. La loro relativa presenza è comunque difficile da stimare, specialmente quando la perturbazione dal sistema integrabile non è piú piccola. Nei sistemi dissipativi, d'altra parte, la proprietà della stabilità asintotica implica la possibilità di *caos attrattivo*. In un sistema che presenta questa proprietà, tutte le traiettorie partenti da una certa parte dello spazio delle fasi ben stabilita e finita, tenderebbero presto o tardi verso l'attrattore; in altri termini, per tali sistemi il caos costituirebbe la regola e si manifesterebbe nel comportamento osservato. La principale questione da porre in riferimento al comportamento caotico nei sistemi dissipativi è dunque se tali sistemi diano origine ad attrattori non appena alcuni parametri di controllo vengono variati.

L'apparire di comportamenti non periodici di tipo turbolento in molti degli esperimenti considerati nel capitolo 1 suggerisce una risposta affermativa a questa domanda. È pertanto importante costruire dei modelli che interpretino le osservazioni e definire condizioni minimali che devono essere soddisfatte per dare inizio al comportamento caotico. Le idee prevalenti fino agli anni '60 suggerivano un quadro disarmante dovuto all'apparire della turbolenza, per lo meno nei fluidi: si pensava alla turbolenza come al manifestarsi della dinamica di una *infinità* di modi accoppiati, ognuno dei quali associato a una particolare frequenza. Inoltre, affinché questi modi fossero del tutto accoppiati, sembravano essere necessarie infinite transizioni oltre alla perdita di stabilità del regolare fluido (laminare) allo stato stazionario. È stato dunque con grande sorpresa ed eccitazione che gli scienziati divennero successivamente consapevoli della possibilità che i sistemi dinamici coinvolgenti solo poche variabili possono presentare un comportamento caotico dopo poche biforcazioni dallo stato di riferimento.

Il primo di tali esempi fu scoperto nel 1963 dal meteorologo americano Edward Lorenz. Il suo modello offre una descrizione semplificata del problema della convezione termica discusso nel capitolo 1, § 3, nel quale si mantiene l'uso di soli tre modi, ognuno associato a una particolare lunghezza d'onda spaziale. Questo dà origine a un attrattore che coglie gli aspetti principali della convezione turbolenta. Un tale fenomeno avviene di routine nell'atmosfera terrestre, ed è all'origine delle ben note difficoltà nella previsione del tempo. Inoltre, ciò suggerisce il fatto che il tempo e il clima sono fondamentalmente impredicibili, dato che dovranno condividere una delle proprietà base del moto caotico, quella di dipendere in maniera sensibile dalle condizioni iniziali. Per sottolineare questo punto molto importante, qualche volta si dice, non del tutto esagerando, che il battito d'ali di una mosca a Cambridge, nel Massachusetts, può benissimo essere all'origine di un grande cambiamento climatico nel continente indiano!

Il lavoro di Lorenz rimase sconosciuto fino agli anni '70, ma da allora c'è stata una vera esplosione di risultati nuovi nella dinamica del caos. Queste idee germogliarono simultaneamente in molte discipline: tra le altre la matematica pura, le simulazioni estensive con il computer, la fisica e la chimica teoriche e sperimentali, la biologia della popolazione. Esse hanno condotto a interessanti equazioni prototipiche che dànno origine a comportamenti caotici, e anche a numerosi «scenari» su come si realizza l'attrattore caotico al variare dei parametri. Daremo ora alcune illustrazioni scelte di ognuno di questi aspetti.

Alcuni prototipi che dànno origine al comportamento caotico.

Un interessantissimo modello matematico del comportamento caotico è stato suggerito da Otto Rössler. Esso contiene tre variabili e solamente una non-linearità di tipo quadratico:

$$\frac{dx}{dt} = -y - z$$

$$\frac{dy}{dt} = -x + ay \tag{3.30}$$

$$\frac{dz}{dt} = bx - cz + xz$$

dove a, b e c sono costanti positive. Le equazioni di evoluzione hanno due punti fissi, uno dei quali è la soluzione banale $x_s = y_s = z_s = 0$, e l'altro è il punto $x_s = c - ab$, $y_s = b - (c/a)$, $z_s = (c/a) - b$. Da qui in avanti discuteremo solamente i fenomeni che avvengono attorno al primo punto fisso, P_0.

Per una larga gamma dei valori del parametro, l'analisi della stabilità lineare considerata nel § 6 e nell'*Appendice* A predice che P_0 può essere instabile. Il comportamento nelle sue vicinanze ha i seguenti aspetti peculiari: le traiettorie vengono respinte lungo una superficie bidimensionale dello spazio delle fasi, nella quale il punto fisso si comporta come un fuoco (figura 3.12*b*); esse sono invece attratte lungo una curva a una dimensione. Chiamiamo tale punto fisso un *fuoco a sella*. Una tale configurazione dà origine a instabilità del moto, un ingrediente base del comportamento caotico, ma permette anche il reinserimento delle traiettorie instabili nelle vicinanze di P_0 e dunque l'eventuale formazione di un attrattore stabile.

Le figure 3.26 *a* e *b* descrivono gli attrattori caotici ottenuti per due diversi valori del parametro. Entrambi gli aspetti appena menzionati sono chiaramente presenti. Si ha però che, mentre in *a* le traiettorie sono reinserite nella stessa zona di P_0 rispetto al piano xy, in *b* alcune di esse sono reinserite nell'altra zona di P_0. Chiameremo questi due moti, rispettivamente, caos di tipo a *spirale* o a *vite*.

Un altro importante aspetto mostrato chiaramente in entrambe le figure 3.26 *a* e *b* è il *piegarsi* della superficie lungo la quale avviene il moto instabile. È istruttivo visualizzarlo sulla superficie della sezione di

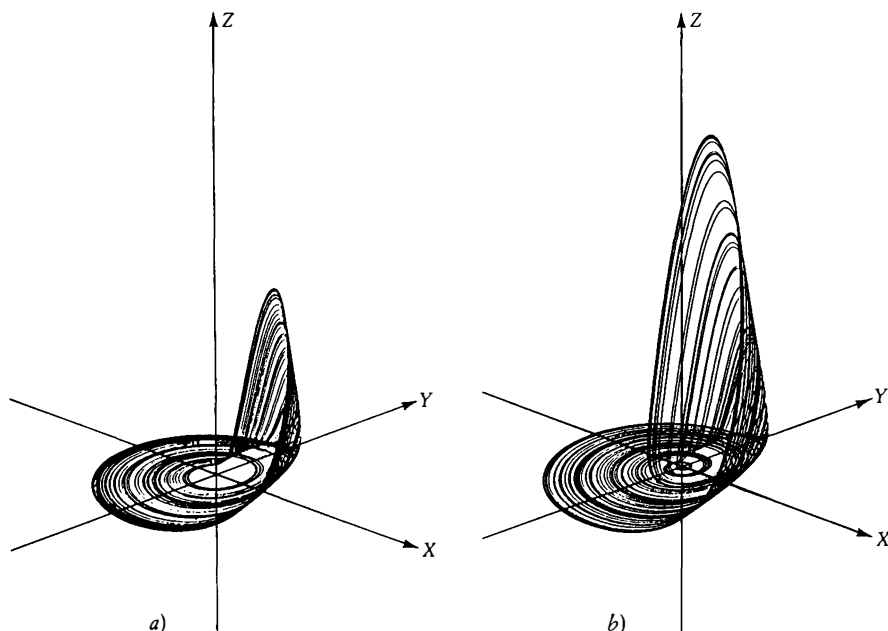
Poincaré ottenuta tagliando trasversalmente il flusso per mezzo del piano $y = 0$, $x<0$, $z<1$. Ciò è illustrato nella figura 3.27. Nella parte *a* si può vedere che l'espandersi delle traiettorie lungo le direzioni instabili e il loro susseguente avvolgersi dà origine a una mappa a ferro di cavallo nel piano di Poincaré. Alla luce dei risultati del § 9, questo spiega qualitativamente il carattere casuale non periodico delle traiettorie sull'attrattore.

Un secondo aspetto di questo fenomeno è mostrato nella figura 3.27*b*. Qui si è disegnato il grafico della posizione del $n + 1$-esimo punto di intersezione tra la superficie di Poincaré sopra definita e il flusso in funzione della posizione all'n-esima intersezione. La costruzione nu-

Figura 3.26

Attrattori caotici ottenuti per integrazione numerica del modello di Rössler (equazione 3.30).

a) Caos a spirale per $a = 0{,}32$, $b = 0{,}3$, $c = 4{,}5$. Le traiettorie vengono immesse sullo stesso lato del punto fisso instabile.

b) Caos a vite per $a = 0{,}38$, $b = 0{,}3$, $c = 4{,}5$. Le traiettorie vengono immesse su entrambi i lati del punto fisso instabile.

merica mostra che si ottiene una curva liscia a forma di campana. Si noti che le posizioni delle intersezioni successive su questa curva non sono date da punti consecutivi ma piuttosto da punti che sembrano essere distribuiti casualmente.

La figura 3.27*b* stabilisce abbastanza naturalmente la connessione tra il nostro modello continuo nel tempo, equazione 3.30, e una seconda classe di sistemi dinamici che mostra il comportamento caotico. Questi sono modelli a tempo discreto definiti dalla seguente equazione iterativa

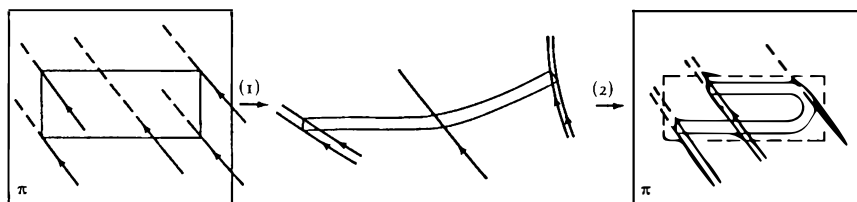
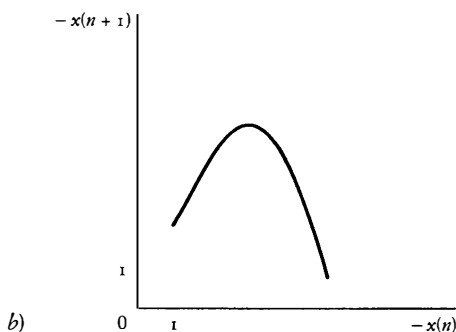
$$x_{n+1} = f(x_n) \tag{3.31a}$$

Figura 3.27

a) Mappa a ferro di cavallo indotta dal sistema dinamico delle equazioni 3.30 sulla superficie di sezione di Poincaré. Il flusso illustrato nella figura 3.26*a* è tagliato dal piano ($y = 0$, $x < 0$, $z = 1$). Le traiettorie che si dipartono dal rettangolo a sinistra della figura ruotano attorno alla varietà stabile unidimensionale del punto fisso, e intersecano ripetutamente la superficie di Poincaré dopo aver seguito il piegarsi della varietà instabile.

b) Mappa unidimensionale ottenuta con il grafico del valore di $-x$ alla (n + 1)esima intersezione della traiettoria con la superficie di Poincaré, in funzione del suo valore alla $n$esima intersezione.

a)

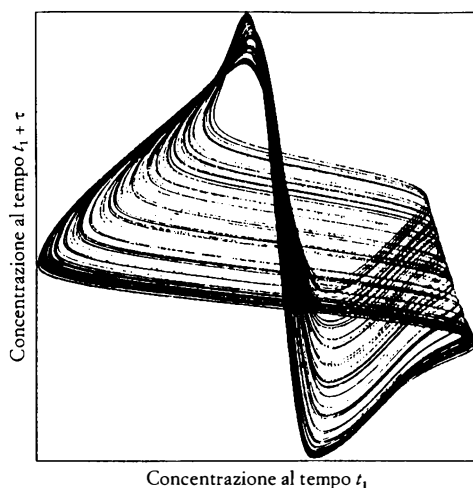
b)

Una tipica curva a forma di campana simile alla figura 3.27*b* si ottiene scegliendo

$$f = 1 - \lambda x^2 \qquad (3.31b)$$

Oltre alla loro connessione con i sistemi continui nel tempo attraverso la mappa di Poincaré, i sistemi di questo tipo possono anche servire direttamente da modello in alcune situazioni in biologia, dove si ha a che fare con popolazioni in cui i gruppi di età degli individui presenti sono determinati dai gruppi di età delle generazioni precedenti.

Nonostante la loro apparente semplicità, le equazioni 3.31 mostrano un comportamento incredibilmente ricco che va da semplici punti fissi a soluzioni multiple periodiche o caotiche. Per di piú, la maggior parte dell'informazione sullo «scenario» per raggiungere il comportamento caotico viene da queste equazioni. Rinviamo dunque ulteriori discussioni al prossimo sottoparagrafo. Chiudiamo la presente rassegna ricordando che esistono numerosi esempi sperimentali che producono attrattori caotici e che dànno luogo, attraverso opportune superfici di

Figura 3.28

Attrattore caotico della reazione BZ in un sistema aperto. Il segnale misurato è proporzionale al logaritmo della concentrazione dello ione Br^-. Gli assi coordinati sono il segnale ai vari tempi t, e il segnale ai tempi $t + \tau$. Il valore del ritardo di tempo scelto per disegnare la figura è $\tau = 53$ sec. (Da J. C. Roux, R. Simoy, e H. Swinney, *Physica*, 8D, 257, 1983).

sezione di Poincaré, a una dinamica temporale discreta simile alla figura 3.27*b*. La figura 3.28 mostra una illustrazione particolarmente sorprendente ottenuta dagli esperimenti sulla reazione di Belusov-Žabotinskij in un reattore ben miscelato.

Alcuni scenari per raggiungere il comportamento caotico.

In contrasto con le biforcazioni statiche e a ciclo limite discusse nei §§ 4 e 5, i meccanismi all'origine degli attrattori caotici non si possono sistemare in una forma universale, normale, dalla quale si possa vedere in modo trasparente l'effetto dei parametri. Piuttosto, combinando qualche generale considerazione con l'esperienza acquisita da esaurienti calcoli numerici, si arriva a regolarità che permettono di identificare certe particolari strade verso il caos. È difficile decidere se questa ineluttabilità sia dovuta alla vera natura del caos, o a un temporaneo inconveniente dovuto alla complessità del problema. La questione è molto simile a quella predominante nella scienza della vita prima dell'avvento della biochimica e della biologia molecolare, dove la tassonomia, ovvero l'arte di classificare in gruppi o in famiglie, era il modello predominante.

Un risultato importante che definisce un particolare scenario del caos è quello dovuto a David Ruelle e Floris Takens. Come abbiamo visto nel § 7, uno dei possibili attrattori nello spazio delle fasi di tre o piú dimensioni è il toro che descrive un moto quasi periodico. Tali tori emergono frequentemente dalle orbite periodiche attraverso biforcazioni secondarie o piú alte. Ruelle e Takens mostrarono che un toro bidimensionale è «robusto» nel senso della stabilità strutturale definita nel § 4; in altre parole, applicando una piccola perturbazione attraverso un cambiamento di parametro si avrà, in generale, un tale attrattore. Non è cosí, tuttavia, per tori di tre o piú dimensioni, che coinvolgono almeno tre frequenze incommensurabili. Questi sono oggetti fragili, e, scomparendo sotto l'azione di una piccola perturbazione, dànno luogo a strani attrattori che mostrano i caratteri discussi dal § 8 in poi e che sono strutturalmente stabili. Tali transizioni sono state osservate su sistemi di modelli e su esperimenti, soprattutto in fluidodinamica.

Un secondo scenario che porta a previsioni piú quantitative, emerge dallo studio della trasformazione temporale discreta definita nell'equazione 3.31. Essa corrisponde a un'infinita sequenza di biforcazioni per valori del parametro ben definiti $\lambda_1 < \lambda_2 < \ldots\ \lambda_n < \ldots$ ognuno dei quali porta a cicli di ordine maggiore il cui periodo raddoppia a ogni succes-

siva biforcazione. I λ_n si accumulano per un particolare punto λ_∞, oltre il quale si ottengono orbite di «periodo infinito» che mostrano un marcato comportamento caotico. Alla fine l'intero spazio dello stato appartiene a un singolo attrattore caotico che mostra instabilità del moto e sensibilità verso le condizioni iniziali.

Un risultato molto importante, dovuto a Mitchell Feigenbaum, è l'esistenza di regolarità nell'approccio alla regione caotica. Esso stabilisce che si raggiunge λ_∞ con una progressione geometrica di rapporto $\delta \simeq 4{,}669$. Inoltre questo comportamento dipende solamente dalla struttura della funzione di iterazione $f(x)$ attorno al suo massimo. È dunque strutturalmente stabile e avviene esattamente con le stesse caratteristiche *quantitative* per intere famiglie di mappe, inclusa la semplice parabola di equazione 3.31, ma anche l'onda sinusoidale, il cerchio, l'ellisse e cosí via. Questa si chiama *universalità della metrica*. La tabella 3.1 illustra la situazione. Il lettore è fortemente incoraggiato a giocare alla «matematica sperimentale» e a controllare da solo la tabella. Sarà sufficiente inserire le equazioni 3.31 in una calcolatrice scientifica tascabile ed eseguire le successive iterazioni.

Molti degli esperimenti esaminati nel capitolo 1 rivelano l'esistenza di procedimenti iterativi («cascate») che raddoppiano il periodo e che portano infine a un comportamento periodico, confermando il valore stimato teoricamente del rapporto di accumulazione δ. Non c'è dubbio

Tabella 3.1

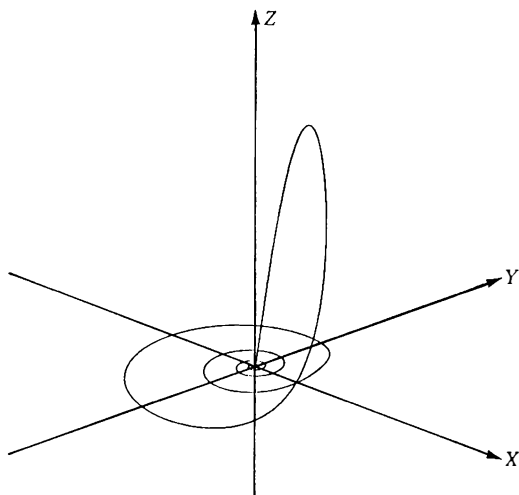
Cascate che raddoppiano il periodo nella mappa $x_{n+1} = 1 - \lambda x_n^2$

n (Orbite periodiche)	Valori di biforcazione	$\lambda_n - \lambda_{n-1}$	$\delta = \frac{\lambda_n - \lambda_{n-1}}{\lambda_{n+1} - \lambda_n}$
0	0.75	–	–
1	1.25	0.5	4.233738275
2	1.3680989394	0.1180989394	4.551506949
3	1.3940461566	0.0259472172	4.645807493
4	1.3996312384	0.0055850823	4.663938185
5	1.4008287424	0.0011975035	4.668103672
.	.	.	.
.	.	.	.
.	.	.	.
∞ (comportamento non-periodico)	1.4011552···	–	4.669···

dunque che la scoperta di questa tendenza sia una conquista fondamentale nel campo della dinamica caotica.

Si può identificare uno scenario ancora diverso per il comportamento caotico nelle trasformazioni temporali discrete, che è noto come l'intermittenza di Pomeau-Manneville. Esso nasce quando una soluzione periodica diventa instabile e dà luogo a traiettorie nelle quali gli intervalli di tempo dei comportamenti praticamente periodici sono seguiti da intervalli di comportamento disordinato sotto forma di scoppi «turbolenti», la cui durata è essa stessa casuale. Questo modello è stato osservato nella maggior parte degli esperimenti esaminati nel capitolo 1. Tornando indietro ai flussi continui nel tempo, la reimmissione delle traiettorie di partenza nella vicinanza di un punto fisso instabile del tipo sella-fuoco (vedere la figura 3.26) è spesso associata alla comparsa di orbite di tipo piuttosto eccezionale, note come *orbite omocliniche*. Queste sono traiettorie che lasciano il punto fisso, ma tornano poi verso di esso: in altre parole, esse tendono allo stesso limite quando il tempo t tende sia a $+\infty$ che a $-\infty$ (si veda la figura 3.29). Esse sono molto sensibili alle variazioni dei valori dei parametri e vengono generalmente distrutte se i parametri non soddisfano una stretta uguaglianza (sono

Figura 3.29

Orbita omoclinica associata al punto fisso (0, 0, 0) del modello di Rössler ottenuta per i seguenti valori dei parametri $a = 0{,}38$, $b = 0{,}30$, $c \simeq 4{,}82$.

strutturalmente instabili nella terminologia del § 4). Tuttavia, per valori vicini dei parametri la loro scomparsa lascia una struttura molto ricca di orbite nello spazio delle fasi, alcune delle quali si comportano in modo caotico. È evidente che alcune delle succitate origini del caos implicano la comparsa di orbite omocliniche. È dunque probabile che lo scenario omoclinico giochi un ruolo sempre piú importante nella dinamica caotica.

Vi sono ancora due questioni riguardanti il comportamento caotico. Primo: è possibile affermare un ruolo maggiore del comportamento caotico rispetto a quello regolare nei sistemi dissipativi che coinvolgono molte variabili? L'esperienza accumulata dallo studio dei sistemi modello e reali suggerisce che entrambi i tipi di comportamento sono possibili e possono anche coesistere. Non c'è pertanto nulla che indichi un'irresistibile tendenza dei sistemi a molte variabili a esplodere nel caos immediatamente dopo la prima biforcazione. Per di piú, per certi tipi di cinetica risulta che localizzare il dominio caotico nello spazio dei parametri è un'impresa complicata. I processi regolatori biochimici forniscono un caratteristico esempio, forse non cosí casuale poiché la riproducibilità deve essere un aspetto essenziale dei ritmi biologici. In quest'ambito è utile ripetere i commenti fatti nel capitolo II, § 4, concernenti ciò che deve essere guardato come «generico» o «tipico». Il comportamento dei sistemi fisici non si riflette necessariamente nei modelli astratti scelti a caso. Ciò è soprattutto ovvio in biologia, in cui ogni funzione presente nelle cellule vive è il risultato di una lunga evoluzione. Abbiamo qui un altro esempio della specificità dei sistemi al non-equilibrio in contrapposizione a quelli all'equilibrio. Infatti, abbassando la temperatura di qualsiasi materiale in condizioni di equilibrio, raggiungiamo alla fine un punto di transizione di fase. Ma se regoliamo a piacere un parametro che interferisce sul comportamento di un sistema al non-equilibrio, osserviamo cicli limite, stati stazionari simili all'equilibrio, caos, o qualcosa di ancora diverso. Possiamo evolvere dal disordine termico dell'equilibrio all'ordine su larga scala, e raggiungere poi quello stato peculiare di disordine che è il caos; ma possiamo anche essere testimoni di un'evoluzione opposta o persino evitare completamente il caos.

La seconda questione concerne il ruolo del caos in natura. Abbiamo già sottolineato il fatto che gli attrattori caotici forniscono l'interpretazione di molti risultati sperimentali. Oltre a questo positivo raffronto, tuttavia, il caos ci rivela un intero nuovo mondo di forme e di modelli. Esso mostra che il disordine in un certo dominio è perfettamente com-

patibile con l'ordine in un diverso dominio, come si può rivelato dalla stessa esistenza di un attrattore. Esso mostra anche che la casualità non è il risultato di imperfezioni sperimentali o di un ambiente complesso che non possiamo controllare, ma, piuttosto, che è profondamente radicata nella dinamica dei sistemi perfettamente deterministici coinvolgenti poche variabili. Soprattutto, la coesistenza di casualità e di ordine, dà luogo a un nuovo concetto che è stato fino ad ora assente nella nostra discussione, cioè il concetto di *informazione*. Analizzeremo brevemente quest'ultimo venuto nel mondo della complessità nel § 13 e, piú in profondità, nel capitolo IV.

11. *Sistemi spazialmente distribuiti. Biforcazioni che rompono la simmetria e morfogenesi.*

Fino ad ora in questo capitolo abbiamo trattato i sistemi dinamici che coinvolgono un numero finito di variabili. Rivolgiamo ora la nostra attenzione ai sistemi spazialmente distribuiti. Qui le variabili di stato (per esempio la temperatura in una piastra soggetta a un riscaldamento sistematico a uno dei suoi bordi) sono campi, nel senso che dipendono continuamente dalle coordinate spaziali. Questo implica dunque l'esistenza di un'infinità di variabili presenti nel problema. Parlando in generale, queste variabili saranno accoppiate, poiché le disomogeneità spaziali daranno luogo a fenomeni di trasporto come la conduzione del calore, la diffusione o la convezione. Ciò verrà riflesso dalla presenza di *derivate spaziali* nelle equazioni di evoluzione, come il laplaciano della temperatura $\nabla^2 T(\mathbf{r}, t)$ o della concentrazione $\nabla^2 c(\mathbf{r}, t)$, il gradiente della velocità $\nabla \mathbf{v}(\mathbf{r}, t)$, e cosí via. Ci aspettiamo che questi sistemi debbano essere molto piú complicati da trattare; tuttavia, in molti casi si possono ottenere interessanti risultati usando opportune procedure perturbative che riducono il problema a un numero di variabili finito.

Questo meccanismo è del tutto simile a quello schematizzato nel § 6 per la riduzione della dinamica di un sistema multivariato a una forma normale. Il punto di partenza è di nuovo l'analisi della stabilità lineare (si veda l'*Appendice* A). In presenza di disomogeneità spaziali, l'operatore di stabilità lineare L che compare nell'equazione 3.25 conterrà derivate spaziali. Ma in molti casi queste derivate compaiono in modo additivo (si vedano ad esempio le equazioni A.14). Cerchiamo allora soluzioni che siano il prodotto di una parte dipendente dal tempo, $c(t)$, e di una funzione $u(\mathbf{r})$ che contenga tutte le dipendenze spaziali (equazione A.19). Ciò porta alla separazione delle variabili, ovvero, a due insiemi

di equazioni piú semplici che coinvolgono solo $u(\boldsymbol{r})$ e $c(t)$. Le equazioni per $u(\boldsymbol{r})$ si possono spesso risolvere esattamente. Per esempio, in un sistema monodimensionale che coinvolge la diffusione, $u(\boldsymbol{r})$ è la sovrapposizione di funzioni seno e coseno: le autofunzioni del laplaciano nel dominio spaziale di cui ci si interessa. La parte dipendente dal tempo $c(t)$ può dunque venire trattata esattamente come nell'equazione 3.25 e se ne può dedurre un grafico simile alla figura 3.16 che descrive la velocità di crescita delle perturbazioni in funzione del parametro di controllo λ. Entra tuttavia in gioco un altro elemento: alcuni dei coefficienti $L_{ij}(\lambda)$ contengono una «traccia» delle disomogeneità spaziali, ovvero l'azione delle derivate spaziali sulla funzione $u(\boldsymbol{r})$ o equivalentemente, l'autovalore della parte dipendente dallo spazio dell'operatore L. Quest'ultimo dipende dal valore dei coefficienti di trasporto e dalla lunghezza spaziale caratteristica del modo descritto da $u(\boldsymbol{r})$.

In un sistema reazione-diffusione questo viene semplicemente dato dai prodotti $-D_i k^2$, dove D_i è il coefficiente di diffusione delle specie i e $k = 2\pi/l$, dove l è la lunghezza spaziale caratteristica (cfr. equazione A.20).

Avendo ora questa libertà addizionale, ci si può porre la seguente domanda. Supponiamo che solo una delle curve di γ in funzione di λ della figura 3.16 attraversi l'asse delle λ. Allo stato marginale $\lambda = \lambda_c$, le caratteristiche del sistema dipenderanno ancora dal parametro addizionale k. È possibile identificare il valore piú «pericoloso» di k che dominerà con ogni probabilità la transizione all'instabilità? Per rispondere a questa domanda è istruttivo fare un grafico del valore del parametro critico λ_c in funzione di k. Due risultati tipici, che corrispondono rispettivamente alle equazioni A.23*a* e *b*, sono descritti nella figura 3.30.

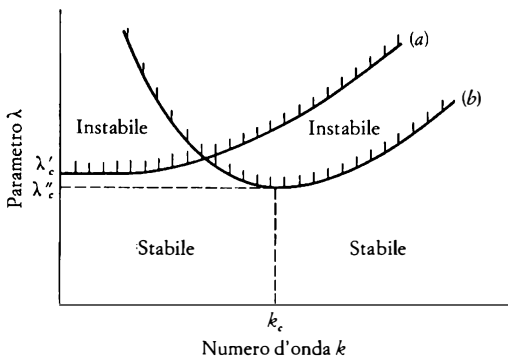
In entrambi i casi *a* e *b*, la curva λ_c in funzione di k ha un estremo, indicati da λ_c' e λ_c'' rispettivamente nella figura. Questo significa che al variare di λ dai valori che corrispondono alla stabilità asintotica (curve inferiori *a* e *b* ai valori che portano all'instabilità (curve superiori *a* e *b*), la *prima* transizione avrà luogo a λ_c' o λ_c'' e dominerà dunque il comportamento del sistema per valori vicini di λ. Ora nel caso *a*, l'estremo λ_c' si ottiene per $k = 0$, cioè, per una caratteristica spaziale $l = \infty$. Questo corrisponde ovviamente a una situazione indipendente dallo spazio. In altre parole, il modo dominante in vicinanza della prima biforcazione sarà un modo omogeneo. L'analisi si ridurrà cosí a quella effettuata nei paragrafi precedenti.

Ma una nuova ed eccitante possibilità sorge nel caso *b* della figura 3.30. Qui l'estremo λ_c'' si realizza per un valore non banale $k = k_c$, cioè, per una ben definita lunghezza spaziale caratteristica $l = l_c$. Ne segue che il modo dominante nelle vicinanze della prima biforcazione sarà spazialmente disomogeneo. Inoltre, le sue caratteristiche saranno *intrinseche*, nel senso che sono interamente determinate dai parametri del sistema. Chiamiamo questa situazione *rottura della simmetria spaziale*.

In entrambi i casi *a* e *b* della figura 3.30 si può ancora definire un opportuno insieme dei parametri d'ordine, che obbediscono a un insieme chiuso di equazioni nella vicinanza della prima biforcazione. La *forma normale* di queste equazioni ha la stessa struttura che nei §§ 4-6 e nell'*Appendice* B. Inoltre, quando parecchi modi diventano simultaneamente instabili, si può estendere la descrizione per mezzo di un numero piú grande di equazioni che descrivono l'interazione fra parecchi parametri d'ordine. Questo dà luogo a biforcazioni secondarie e di ordine piú alto, che portano, oltre ad altri fenomeni, la coesistenza di molte strutture spaziali stabili. Infine, in sistemi di ampia estensione spaziale, il numero di modi interagenti tende all'infinito con il risultato che le normali equazioni di forma per i parametri d'ordine divengono equa-

Figura 3.30

Due tipici risultati della stabilità lineare attorno a uno stato stazionario uniforme di un sistema spazialmente distribuito.

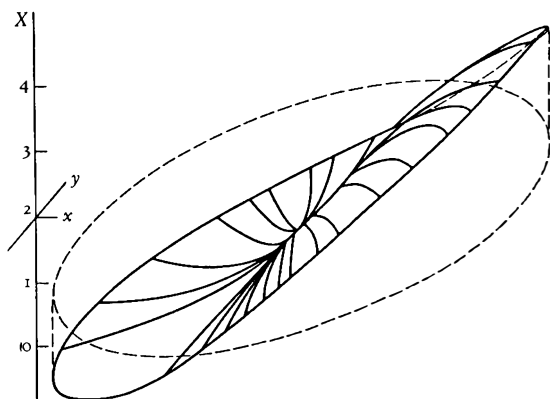
La curva di stabilità marginale *a* ha un estremo per il valore del numero d'onda $k = 0$. Su queste basi ci si aspetta che il modo dominante oltre l'instabilità rimarrà spazialmente omogeneo; la curva di stabilità marginale *b*) ha un estremo per $k = k_c \neq 0$. Ci si aspetta che i rami della biforcazione emergenti dopo la transizione mostrino una dipendenza spaziale non banale. Questa è la base del fenomeno della rottura di simmetria spaziale.

zioni differenziali parziali. È comunque ancora possibile dimostrare che anche per tali sistemi le traiettorie tendono a un attrattore finito dimensionale immerso in uno spazio delle fasi di dimensione finita, detto *varietà inerziale*. Queste dimensioni possono diventare alte se il sistema opera lontano dalla soglia dell'instabilità. L'inizio della struttura spaziale in un mezzo fino a quel momento omogeneo, è un problema centrale in molti campi. In idrodinamica, le strutture spaziali accompagnano frequentemente la transizione all'instabilità, come nel problema di Bénard descritto nel capitolo I. Nella scienza delle interfacce e dei materiali esse dànno luogo a forme affascinanti delle quali le dendriti sono un esempio sorprendente. Nello sviluppo embrionale le strutture spaziali forniscono la necessaria «informazione posizionale» per la differenziazione delle cellule, come discusso brevemente nel capitolo II, § 7. In quest'ultimo caso – e in effetti in tutti i problemi che coinvolgono un accoppiamento del tipo diffusione fra le variabili – un aspetto interessante è l'esistenza di una *grandezza minima del sistema* sotto la quale le strutture spaziali non si possono formare. Intuitivamente, questo si può facilmente comprendere. In un sistema reattivo spazialmente distribuito, le reazioni chimiche e la diffusione cooperano a una biforcazione che rompe la simmetria nel modo seguente. La cinetica chimica, attraverso il suo intrinseco *feedback*, può dar luogo localmente a un fenomeno di scatenamento amplificando l'effetto di piccole perturbazioni o fluttuazioni. La diffusione tende a disperdere le disomogeneità, ma quando la sua velocità è confrontabile con la velocità di reazione, essa non vi riesce. Il risultato è allora una struttura spaziale, la cui lunghezza caratteristica corrisponde alla distanza media sopra la quale un gruppo di molecole reattive si può diffondere prima che si verifichi una reazione. Ora, in un sistema di piccole dimensioni, le reazioni avvengono alla stessa velocità che in un sistema a grandi dimensioni, ma la diffusione è tremendamente accresciuta (il tasso di diffusione è proporzionale a D/l^2, indicando con D il coefficiente di diffusione e l le dimensioni del sistema). Il sistema viene dominato dal «rimescolamento» da essa imposto, e le condizioni al contorno distruggono ogni tentativo locale di destabilizzazione.

Le strutture spaziali che sorgono attraverso le biforcazioni dovute a rottura di simmetria, possono essere stazionarie come anche dipendenti dal tempo. Un esempio di strutture stazionarie è dato dalla figura 3.31. Un insieme di due equazioni di reazione-diffusione accoppiate è stato numericamente simulato sopra un disco circolare. Vediamo chiaramen-

te la comparsa di *assi di polarità* nel sistema, che inducono una differenziazione lungo una direzione privilegiata che in un esperimento reale può essere un asse verticale o dorsoventrale. Sorprendenti esempi sperimentali di strutture spaziali dipendenti dal tempo sono le onde osservate nei reagenti della BZ e nel *Dictyostelium discoideum*, mostrati nelle figure 1.9 e 1.15. Tali strutture si possono ottenere anche da sistemi modello. Nella figura 3.32 vengono dati due esempi che dànno luogo, rispettivamente, a semplici onde rotanti e a spirali rotanti.

Una proprietà caratteristica delle strutture temporali sia stazionarie che dipendenti dal tempo, alla quale abbiamo già alluso nel cap. II, § 7, è quella di apparire sotto forma di *multipletti* i cui membri hanno struttura e proprietà di stabilità identiche in un mezzo simmetrico. Per esempio nella figura 3.31 l'asse di polarità potrebbe essere un qualsiasi diametro del cerchio. Analogamente, in un mezzo oscillante infinitamente esteso, le spirali appaiono a coppie, i due membri delle quali ruotano in direzioni opposte (in senso orario e antiorario). Cosí, ogni spirale è una struttura asimmetrica che mostra una privilegiata *chiralità*, ma in una popolazione di spirali di un sistema macroscopico la simmetria viene ristabilita dalla coesistenza di numeri uguali di forme di opposto senso di rotazione. Il problema della selezione fra tali forme verrà discusso nel § 13.

Figura 3.31

Veduta in prospettiva di una struttura di concentrazione spaziale X generata su un dominio circolare (sul piano $x-y$). La struttura è il prodotto di una parte radiale, la funzione di Bessel di primo ordine $J_1(r)$, e una parte angolare, cos φ.

Un'affascinante classe di strutture dipendenti dallo spazio e dal tempo nasce quando la diffusione, o qualche altro fenomeno di trasporto, induce una dinamica caotica in un sistema omogeneo che normalmente si comporta in modo regolare. La figura 3.33 illustra una tipica configurazione ottenuta da simulazioni numeriche sul modello del brussellatore. Si osservano comportamenti analoghi negli oscillatori accoppiati dove, a seconda della forza dell'accoppiamento, il sistema può essere portato da un regime di perfetto accordo di fase a un regime caotico caratterizzato da una marcata asincronizzazione della fase.

12. *Sistemi dinamici discreti. Automi cellulari.*

In molti casi un fenomeno fisico implica una soglia ben precisa. Oltre tale soglia la velocità di un processo si stabilizza a qualche valore che è completamente differente dal valore prevalente sotto la soglia, e molto spesso l'intervallo di transizione fra questi valori è molto piccolo. Per esempio, un circuito elettronico che include un elemento attivo come un transistor varierà fra due livelli di corrente su un intervallo molto breve di valori di tensione. Tali circuiti sono i costituenti fondamentali dei computer. In biologia, un gene dirige la sintesi di un prodotto, diciamo una proteina, in modo tale che la concentrazione di quest'ultima è praticamente zero quando il gene è inattivo, fatto che può a sua volta essere influenzato dal livello di concentrazione di un effettore accoppiato con l'attività della proteina, ecc. In questo paragrafo descriviamo brevemente il comportamento dinamico di sistemi i cui elementi presentano fenomeni di soglia.

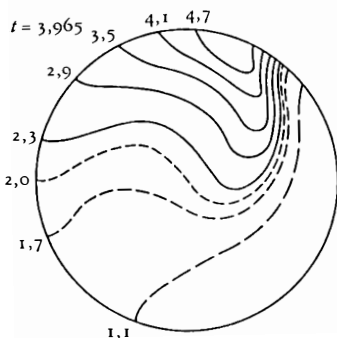
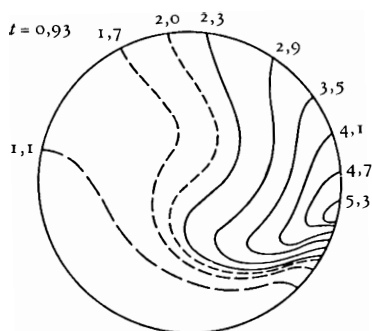
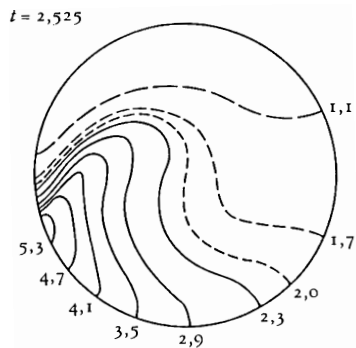
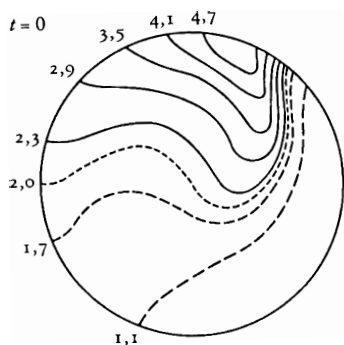
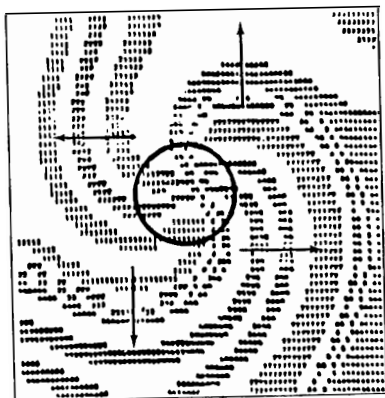
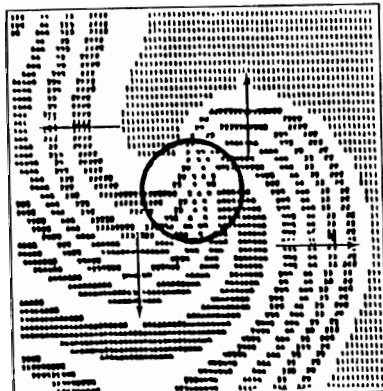
Consideriamo dapprima un semplice elemento che coinvolge solo una variabile, X. La seguente è una tipica situazione: per $0<X<X_0$ la

---

Figura 3.32

a) Onde rotanti ottenute con la soluzione numerica del modello del brussellatore (equazione 3.19) in un cerchio di raggio 0,5861 soggetto alle condizioni di flusso zero al contorno. Le linee piene e tratteggiate si riferiscono rispettivamente, a concentrazioni piú grandi o piú piccole dei valori allo stato stazionario (instabile). Valori dei parametri: $A = 2$, $D_1 = 0{,}008$, $D_2 = 0{,}004$, $B = 5{,}8$.
b) Spirali ruotanti ottenute da soluzioni numeriche del modello

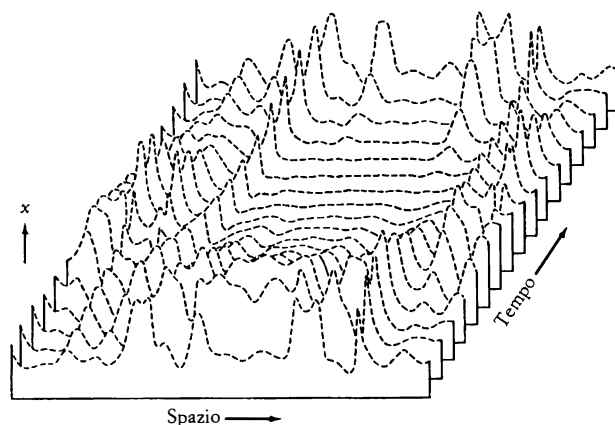
$$\partial A/\partial t = \nabla^2 A - A - B + (A \text{ se } A > 1/20$$
$$\partial B/\partial t = \nabla^2 B + A/2$$

(Da A. Winfree, *SIAM-AMS Proceedings*, 8, 13 1974)

rapidità di cambiamento $F(X)$ è molto piccola, mentre per $X > X_0$, essa diventa immediatamente apprezzabile e si satura subito dopo per qualche valore F_{max}; questo viene mostrato nella figura 3.34. Per cogliere l'essenza di questa risposta guardiamo F come una funzione discontinua della variabile discontinua X. In altre parole, assumiamo che in $0 < X < X_0$ sia X che F siano «zero», mentre per $X > X_0$, X e F siano uguali a «uno». Tali quantità sono note come *variabili booleane*, dal matematico inglese del XIX secolo George Boole, che viene considerato il fondatore della logica formale.

L'evoluzione temporale di X verrà dettata da opportune equazioni logiche che legano X e F e che sostituiscono le equazioni differenziali scritte nel formalismo continuo tradizionale (si veda, ad esempio, l'equazione 3.1). L'esempio piú semplice è dato da:

$$X_{t+1} = F(X_t) \qquad (3.32)$$

in cui le variabili presenti stanno cambiando valori in piena sincronia e in passi temporali discreti. La forma di F dipende dal problema specifico considerato. Per esempio, supponiamo che sia coinvolto un ciclo positivo di *feedback* (un esempio è il *feedback* dell'albedo del ghiaccio discussa nel capitolo I, § 8). In termini logici questo significa che se X (la

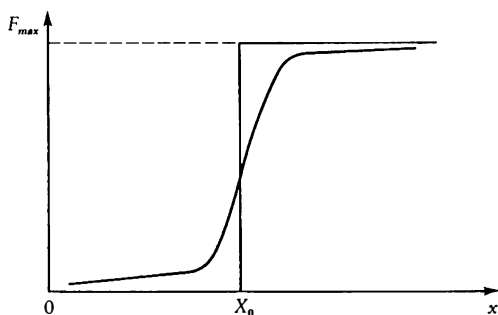
Figura 3.33
Rappresentazione schematica del caos indotto dalla diffusione ottenuta dalla soluzione numerica del modello del brussellatore. (Da Y. Kuramoto, *Physica*, 106A, 128, 1981).

temperatura nell'esempio dell'albedo) è alta la dinamica tenderà ad aumentarla ulteriormente; in altre parole, il valore di $F(X)$ sarà anch'esso alto, e viceversa se X è basso. Ricordando che F e X sono variabili booleane, possiamo esprimere questa dipendenza con l'equazione logica $F = T$[1]. Le soluzioni di questa equazione sono ben visualizzate in una tabella di stati in cui vengono successivamente inseriti i valori logici di F e X:

F	X
0	0
1	1

Vediamo che esistono due stati, uno (0, 0) in cui F e X spariscono simultaneamente, e uno (1, 1) in cui sono simultaneamente presenti. Tali stati non possono evolvere ulteriormente nel tempo e sono l'analogo dei punti fissi dei sistemi dinamici continui. Abbiamo dunque un modello molto semplice della capacità dei sistemi con *feedback* positivo di soggiacere a transizioni fra diversi stati stazionari. Situazioni piú complicate che coinvolgono molte variabili, e anche *feedback* sia positivi che negativi, si possono analizzare direttamente.

Una situazione molto interessante nasce dai sistemi che consistono in un reticolo di componenti booleane identiche che interagiscono lo-

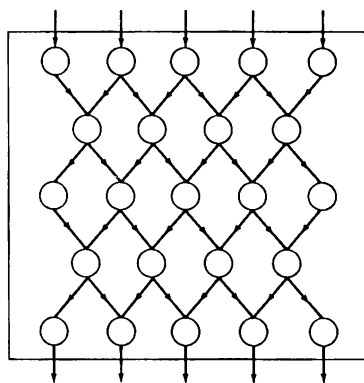
[1] Si noti che nonostante la sua apparente forma lineare questa equazione esprime in realtà la dipendenza altamente non lineare mostrata in fig. 3.34.

Figura 3.34
Rappresentazione di una funzione booleana F, di una variabile booleana X.

calmente nello spazio. Tali sistemi dinamici vengono chiamati *automi cellulari*. Un tipico automa cellulare monodimensionale è descritto dalla

$$X_{i,t+1} = F_i(X_{i-r,t}, X_{i-r+1,t}, \ldots, X_{i,t}, \ldots, X_{i+r,t}) \tag{3.33}$$

che stabilisce che la variabile (booleana) X di un particolare posto i viene determinata dai valori precedenti di un intorno di $2r + 1$ posti attorno a essa. A seconda delle regole che intercorrono in F_i, i sistemi dinamici di questo tipo e i loro analoghi bidimensionali possono generare una varietà di strutture spaziali spettacolari molto simili a quelle descritte nel § 11.

Ci poniamo ora un diverso tipo di domanda. Un insieme di «processori» agisce su dati (booleani) ricevuti localmente dai suoi vicini, come per esempio nella figura 3.35. Ogni elemento, il cui stato interno viene anche rappresentato da una funzione booleana, «calcola» il suo *output*, che viene trasmesso ai suoi vicini. Può una rete di questo tipo dar luogo a strutture di «calcolo» globali come una funzione del tempo che rappresenti qualche nuova *proprietà emergente* non inclusa nelle «programmate» unità iniziali? In linea di principio, la risposta a questa domanda è sí. È basata sull'esistenza di attrattori del sottostante sistema dinamico (dissipativo). Un attrattore corrispondente a una struttura che si autoorganizza in seguito a una instabilità dovuta a rottura di simmetria, necessariamente dota il sistema di una proprietà collettiva che trascen-

Figura 3.35
Una serie di processori, ognuno dei quali opera su dati ricevuti localmente dal suo vicino.

de le proprietà delle sottounità individuali. In realtà, in un sistema di grande estensione spaziale vi saranno molti attrattori. A seconda dello stimolo iniziale, il sistema finirà in diversi bacini di attrazione. Ciò permette di prevedere la possibilità di macchine che impareranno esattamente come riconoscere differenti *input*, anche quando gli *input* sono leggermente distorti.

Il ruolo degli attrattori nel calcolo – o nel pensiero, se è per questo – dà luogo a numerose conclusioni affascinanti nell'interfaccia fra la teoria dei sistemi dinamici e la fisica del non-equilibrio da una parte, e la scienza cognitiva e dei computer da un'altra parte. È probabile che i risultati che scaturiranno da questa inaspettata sintesi porteranno molto lontano.

13. *Asimmetria, selezione e informazione.*

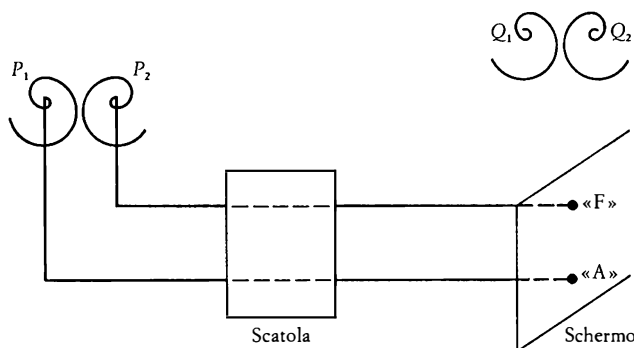
Abbiamo osservato che le soluzioni spazialmente asimmetriche che nascono dalla biforcazione in un mezzo inizialmente perfettamente simmetrico hanno uno speciale status: in una particolare situazione fisica solo una tale soluzione viene realizzata in una regione di spazio sufficientemente piccola. Per esempio, abbiamo visto nella figura 1.9, una particolare parte del reagente della BZ diventa un centro dal quale si emana solo una spirale di una data chiralità, per esempio sinistrorsa. In questo senso dunque, si può parlare di *rottura di simmetria*. Tuttavia, se si guarda alla stessa parte del sistema in un grande numero di esperimenti non correlati, si osserveranno soluzioni di opposte simmetrie in uguale frequenza. Analogamente, in un dato esperimento condotto in un sistema di ampia estensione spaziale, le soluzioni con opposta simmetria coesisteranno nelle diverse parti del sistema. In questo senso, dunque, si ristabilisce la simmetria del mezzo iniziale.

Nelle usuali condizioni di laboratorio vi è da aspettarsi la conclusione precedente. Dopo tutto, non si vede perché una cella di Bénard che ruota, per esempio, in senso orario anziché antiorario, dovrebbe apparire di preferenza in un posto particolare dell'apparato sperimentale. Analogamente, gli esperimenti con i reagenti della BZ hanno sempre mostrato che le onde a spirale appaiono a coppie, a meno che lo sperimentatore non si sia preso cura di favorire una particolare configurazione.

In molti altri casi, l'incapacità di selezionare un particolare stato asimmetrico della materia fra la moltitudine delle forme disponibili

porta a qualche problema fondamentale. Per capire la natura della questione, eseguiamo il seguente esperimento ideale, riassunto nella figura 3.36. Supponiamo che in un appropriato reagente chimico si realizzi uno stato di onde a spirale rotanti. Siano (P_1, Q_1, ...),(P_2, Q_2, ...) i centri delle coppie di spirali destrogire e levogire che si incontrano quando ci si muove lungo il sistema. Supponiamo inoltre che tramite un opportuno disegno del circuito la spirale levogira possa generare, su uno schermo, un simbolo, per esempio «F», mentre tramite un simile circuito, la spirale destrogira possa generare il simbolo «A». Un possibile uso di questi simboli è quello che segue: quando lo schermo mostra il segnale emesso da P_2 sopra a quello di P_1, la struttura risultante (leggendo dall'alto al basso) è un'istruzione per eseguire un compito: «FA». Si assume naturalmente che il «testo» costruito da questo algoritmo deve poter essere letto e compreso dal «lettore» a cui è stato assegnato questo compito. Nell'esempio scelto, questo si verifica se il nostro lettore possiede una rudimentale conoscenza dell'italiano.

Seguiamo il funzionamento di questo meccanismo quando il reagente viene portato al regime delle soluzioni di onde rotanti. Scegliamo i punti P_1 e P_2 (figura 3.36) in modo tale che nel primo esperimento eseguito P_1 sia il centro di una spirale destrogira e P_2 il centro di una spirale

Figura 3.36
Una «macchina di biforcazione». Dotando un sistema con una molteplicità di soluzioni, la biforcazione crea potenzialmente informazione. Un simbolo «F» o «A» viene emesso a seconda se ha luogo una biforcazione di onde a spirale levogira o destrogira. In assenza di selezione il verso di rotazione della particolare onda che sarà realizzata nelle successive operazioni dell'apparato sperimentale non può essere riprodotto. Ossia non v'è alcun modo per assicurare che un'identica coppia Q_1, Q_2 seguirà P_1, P_2. In questo caso la successione dei simboli «F» e «A» avverrà in modo casuale, e questo comprometterà l'elaborazione dell'informazione.

levogira. L'istruzione «FA» apparirà immediatamente sullo schermo, verrà interpretata adeguatamente, e il lettore eseguirà il compito assegnato. La macchina verrà spenta dopo un po' di tempo che funziona con successo. Vediamo che cosa accadrà quando verrà utilizzata nuovamente.

A causa della molteplicità dei regimi possibili, i punti P_1 e P_2 possono ora diventare i centri delle spirali sia levogire che destrogire. Cosí, la nostra macchina non genererà necessariamente lo stesso messaggio «FA» come prima. Invece c'è una probabilità molto grande che il nostro lettore veda messaggi che non può comprendere, come per esempio, «AF», «AA», «FF». Il risultato sarà che egli non eseguirà il compito che gli è stato assegnato. In altre parole, la «macchina della biforcazione» che abbiamo progettato si rivela essere un *elaboratore di informazione* inattendibile al fine di completare il compito assegnato.

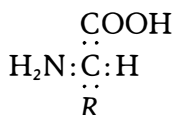
Supponiamo tuttavia di adottare un trucco. Dopo il primo funzionamento della macchina, installiamo due piccoli motori sotto i punti P_1 e P_2, che mescolano debolmente il mezzo rispettivamente verso destra e verso sinistra. Grazie a questo accorgimento verrà invariabilmente selezionata un'onda rotante a destra e a sinistra rispettivamente nelle vicinanze di P_1 e P_2, e ne risulta che lo stesso messaggio «FA» verrà trasmesso ogni volta che il sistema verrà messo in funzione. Ovviamente ora abbiamo una macchina affidabile.

I motori che abbiamo introdotto nel nostro semplice esempio non sono che un esempio del concetto piú generale di *selezione*. In altre parole, la rottura della simmetria spaziale è il prerequisito necessario senza il quale semplicemente non sorgerebbe il problema di costruire un elaboratore di informazione. Ma, infine, è la selezione che mette in grado di scoprire, interpretare, e trasmettere il «messaggio» nascosto nella dinamica non-lineare di non-equilibrio. *La selezione decodifica l'informazione* e, attraverso essa, permette il trasferimento della complessità da un livello a un altro.

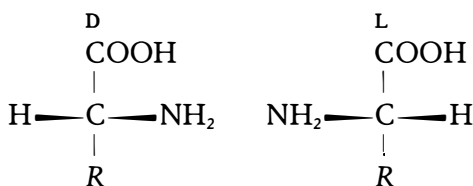
Ma il bisogno di selezione va al di là del nostro semplice esempio. Nel capitolo I, § 9, abbiamo menzionato la rottura di simmetria fra materia e antimateria. Proprio perché nel presente universo la materia è stata «selezionata» rispetto all'antimateria i nuclei, le galassie o gli esseri viventi si sono potuti formare in luogo di una progressiva annichilazione materia-antimateria che produce solo fotoni. Piú vicino a noi, la selezione di forme privilegiate asimmetriche è alla base della struttura e del funzionamento degli organismi viventi di oggi. Esaminiamo pochi

esempi caratteristici che mostrano l'onnipresenza dell'asimmetria in questo campo.

Sappiamo dalla biologia elementare che le proteine sono fra i costituenti basilari della vita. Indipendentemente dalla loro origine e funzione, tutta la sorprendente varietà di proteine esistenti sorge dal combinarsi delle stesse 20 specie di *amminoacidi* in diverse proporzioni molari. Gli amminoacidi trovati nelle proteine hanno quattro gruppi collegati a un atomo centrale di carbonio, detto carbonio α:

$$\begin{array}{c} \mathrm{COOH} \\ \cdot\cdot \\ \mathrm{H_2N : \underset{\cdot\cdot}{C} : H} \\ R \end{array}$$

I quattro gruppi sono:

1. un carbossile, abbreviato COOH
2. un gruppo amminico o NH_2
3. un atomo di idrogeno, H
4. un gruppo laterale o gruppo R diverso per ogni specie.

In tutti gli amminoacidi tranne la glicina, per la quale R è identico a H, il carbonio α è *asimmetrico* perché ha diversi gruppi attaccati a esso. Ci sono quindi due possibili configurazioni di amminoacidi, gli isomeri D e gli isomeri L:

Una coppia di isomeri D e L non si può far coincidere per rotazioni. Essi sono come una mano destra e sinistra e sono analoghi sotto questo aspetto alla coppia di spirali generata nei reagenti della BZ. La loro asimmetria si riflette nella loro abilità di ruotare il piano di polarizzazione della luce polarizzata lungo una direzione privilegiata, quella destrogira per gli isomeri D e levogira per gli isomeri L.

Nelle proteine, gli amminoacidi si trovano sempre nella configurazione L. Abbiamo dunque un esempio sorprendente di una assoluta se-

lezione di asimmetria. Il grande scienziato francese Louis Pasteur, fondatore della moderna biochimica, guardava questo fenomeno come a uno degli aspetti fondamentali della vita.

Ma siamo lontani dall'esaurire la lista di esempi di asimmetria in biologia. Gli acidi nucleici, gli altri costituenti fondamentali della vita oltre alle proteine, sono composti da sottounità note come nucleotidi. In aggiunta a una parte consistente di una base purinica o piridinica – essenzialmente una molecola planare – essi contengono uno zucchero che mostra una struttura chirale tridimensionale: il D-deossiribosio nel caso dell'acido deossiribonucleico (DNA), e il D-ribosio nell'acido ribonucleico (RNA). Ancora, tutti gli acidi nucleici di organismi viventi sembrano essere il risultato di una selezione assoluta in cui si trova solo l'isomero destrogiro D.

Un caso di asimmetria privilegiata ancora diverso, ma ugualmente sorprendente, si trova nel modo in cui viene letto il codice genetico. È acquisito che il codice è un codice a tripletta, in cui ogni simbolo corrisponde a un insieme di tre nucleotidi (*codons*). Per di piú, esistono dei *segnali di partenza* che permettono al messaggio genetico di venire «letto», a tre nucleotidi alla volta, in una *direzione fissata* dal punto di partenza. L'importanza di questa asimmetria è primordiale, poiché la sua assenza produrrebbe effetti analoghi a quelli di una lettura di un testo indifferentemente da sinistra a destra o da destra a sinistra. A questo proposito vale la pena di notare che i nucleotidi sintetici, che non posseggono questa polarità, hanno capacità di informazione molto ridotte.

A questo punto dovremmo esserci convinti del ruolo costruttivo della selezione. È dunque importante indagare sui meccanismi dai quali essa può prendere l'avvio.

Il caso piú evidente di selezione nasce quando la biforcazione con rottura di simmetria è un evento unico, non ripetibile. Un esempio è la selezione della materia rispetto all'antimateria, che, una volta fatta, porta l'universo a un «punto di non ritorno» oltre il quale è impossibile immaginare una realizzazione diversa; almeno per un ordine temporale di 20 miliardi di anni! Un caso simile si ottiene se la transizione alla forma asimmetrica non è spontanea, ma coinvolge uno stadio di nucleazione in cui il sistema deve prima essere «attivato» e superare una «barriera». Se la quantità di «energia di attivazione» richiesta è molto grande, la transizione a un particolare stato asimmetrico è molto lenta. È quindi concepibile che al momento in cui sta per nascere lo stato di opposta simmetria, il primo abbia preso il sopravvento, abbia modificato l'am-

biente, e inibito la successiva comparsa di ogni forma di simmetria opposta.

Un secondo meccanismo generale di selezione è il trasferimento di una asimmetria esterna al particolare sistema in considerazione. Questo solleva immediatamente due questioni. Primo, questa «primordiale» asimmetria deve sempre agire nello stesso modo. Un esempio in cui questa condizione è soddisfatta è la gravità, che influenza inevitabilmente ogni esperimento terrestre coinvolgente oggetti massicci, poiché introduce la verticale come direzione privilegiata. Inoltre, tuttavia, il sistema deve essere sufficientemente sensibile per percepire questi effetti, che sono di solito molto deboli. È qui che le biforcazioni giocano di nuovo un ruolo molto importante. Invero, come abbiamo visto nel § 4, la biforcazione viene profondamente influenzata da piccoli cambiamenti dei parametri che appaiono nelle equazioni di evoluzione del sistema e può anche venire distrutta sotto certe condizioni (cfr., ad esempio, la figura 3.10*c*). Se, dunque, durante la sua storia, a un sistema accade di rimanere sufficientemente vicino a un punto di biforcazione per un lungo periodo di tempo, esso diventerà sensibile ai piccoli effetti che possono concepibilmente indurre la selezione di una particolare forma di asimmetria. Discuteremo ulteriormente questo meccanismo nel capitolo IV, in cui vi vengono sviluppati i concetti necessari per trattare le perturbazioni stocastiche.

In entrambi i meccanismi di cui sopra è presente l'irreversibilità, che assicura, come ripetutamente messo in evidenza, la stabilità asintotica dei differenti stati. Si può intravedere, tuttavia, un meccanismo di selezione ancora diverso in cui essa gioca un ruolo ancora piú diretto. In particolare, si può sollevare la questione se l'universale rottura di simmetria per inversione temporale associata all'irreversibilità, può essere essa stessa una sorgente di selezione che porta a forme di materia ricche di informazioni e *spazialmente* asimmetriche. A questo punto non siamo in grado di poter trattare questa questione. Dobbiamo prima definire in modo appropriato la nozione di informazione, che fino ad ora è stata introdotta piuttosto approssimativamente, usando il linguaggio dell'intuizione. Si arriverà a questo solo dopo che verrà esposta la teoria stocastica dei sistemi dinamici, alla quale rivolgiamo ora la nostra attenzione.

Capitolo quarto

Aspetti stocastici dei fenomeni complessi

Abbiamo precedentemente accennato al ruolo che hanno gli elementi di casualità nell'emergere del comportamento complesso. In tutti i sistemi fisico-chimici questi elementi sono continuamente presenti attraverso il meccanismo delle *fluttuazioni termodinamiche* interne. Inoltre, in molte circostanze, anche l'ambiente che interferisce con il sistema è turbolento; ad esempio un ecosistema che sia soggetto al variare delle stagioni o ad altre fluttuazioni climatiche. In questo capitolo tratteremo la dinamica delle fluttuazioni. Porremo particolare enfasi sui meccanismi per i quali esse interferiscono con il comportamento macroscopico, e sui nuovi concetti che determina un approccio probabilistico ai sistemi dinamici. Risulterà che gli effetti stocastici giocheranno un ruolo cruciale nelle vicinanze dei punti di biforcazione.

1. *Fluttuazioni e descrizione probabilistica.*

L'osservazione dello stato di un sistema fisico-chimico coinvolge, di regola, una media dei valori istantanei delle variabili pertinenti, sia rispetto al tempo sia rispetto a un volume di dimensioni supermolecolari. Ad esempio, se noi mettiamo circa $0{,}33 \times 10^{23}$ molecole di acqua in un contenitore di un centimetro cubo di volume a temperatura e pressione ambiente, concluderemmo che abbiamo ottenuto un liquido la cui densità numerica è di $0{,}33 \times 10^{23}$ molecole per cm^3 e la cui densità di massa è pari a $0{,}33 \times 10^{23} \times$ (massa delle molecole di H_2O) = 1 g/cm^3. Questo è il tipo di variabili con il quale abbiamo avuto a che fare finora in gran parte dei precedenti capitoli. D'altra parte, il valore della densità in piccoli elementi di volume dell'ordine, diciamo, di dieci angstrom cubici (10 $Å^3$), sarà continuamente soggetto a deviazioni dal valore macroscopico. Molecole attraverseranno le frontiere di questi volumi e, per il carattere casuale dei loro moti, sul quale già insistemmo all'inizio (cfr. capitolo 1, § 1), il numero di particelle contenute in qualunque istante

in ogni piccolo volume sarà essenzialmente impredicibile. Chiamiamo *fluttuazioni* le deviazioni generate da questo meccanismo. Grazie alle fluttuazioni, i sistemi fisico-chimici diventano quindi capaci di esplorare in modo continuo lo spazio delle fasi e di rappresentare escursioni attorno allo stato predetto dalla soluzione delle equazioni fenomenologiche e deterministiche.

L'approccio naturale al problema delle fluttuazioni è in termini della teoria di probabilità. Per esempio, in un sistema soggetto a reazioni chimiche le quantità interessanti sono le probabilità, $P(X_\alpha, \Delta V, t)$, di avere, al tempo t, X_α particelle di tipo α nell'elemento di volume ΔV. Un'informazione piú completa sarà contenuta nelle probabilità multivariate, $P(X_{\alpha j}, \Delta V_j; X_{\beta k}, \Delta V_k; \ldots, t)$, di avere simultaneamente occupati i vari elementi di volume j, k, ..., dalle specie chimiche di tipo α, β ...

Un modo molto conveniente di caratterizzare una distribuzione di probabilità è quello di studiarne i *momenti*, ovvero l'insieme delle seguenti quantità

$$(4.1) \qquad \langle X_{\alpha j}^{m_{\alpha j}} \cdot X_{\beta k}^{m_{\beta k}} \cdots \rangle = \sum_{X_{\alpha j}, X_{\beta k} \cdots} X_{\alpha_j}^{m_{\alpha j}} X_{\beta_k}^{m_{\beta k}} \cdots P(X_{\alpha j}, \Delta V_j; X_{\beta k}, \Delta V_k; \ldots)$$

con $m_{\alpha j}$, ecc. interi. Due momenti particolarmente interessanti sono il valor medio, $\langle X_{\alpha j}\rangle$ e il secondo momento $\langle X_{\alpha j} X_{\beta k}\rangle$. Di fatto in molti casi è preferibile ragionare in termini di una quantità di eccedenza nota come *covarianza*,

$$(4.2) \qquad \langle \delta X_{\alpha j} \delta X_{\beta k}\rangle = \langle X_{\alpha j} X_{\beta k}\rangle - \langle X_{\alpha j}\rangle \langle X_{\beta k}\rangle$$

Illustriamo ora il significato di queste quantità nel caso particolare di una singola variabile stocastica, X, per il quale l'espressione 4.2 si riduce alla *varianza*, $\langle \delta X^2\rangle$, della sottostante distribuzione di probabilità. Prima introduciamo una misura di «dispersione statistica» attraverso la *deviazione standard*

$$(4.3) \qquad \delta = (\langle \delta X^2\rangle / \langle X\rangle^2)^{1/2}$$

Un piccolo δ implica una distribuzione con un picco molto acuto attorno al suo valor medio $\langle X\rangle$, che a sua volta è vicino al valore piú probabile, $\bar{X}$, per il quale $P(X)$ assume il suo unico massimo. Un esempio estremo di questa situazione è dato da una distribuzione di probabilità con un picco di forma infinitamente acuta, $P(X) = \delta^{kr}_{X, \bar{X}}$ cioè

$$P(X) = 1 \qquad \text{se } X = \bar{X}$$
$$= 0 \qquad \text{se } X \neq \bar{X} \tag{4.4}$$

In altre parole, tutti gli stati eccetto uno in particolare, $X = \bar{X}$, non sono occupati. La media e la varianza sono ovviamente dati da

$$\langle X \rangle = \sum_X X\, \delta^{kr}_{X,\bar{X}} = \bar{X}$$

$$\langle X^2 \rangle = \sum_X X^2\, \delta^{kr}_{X,\bar{X}} = \bar{X}^2$$

e

$$\langle \delta X^2 \rangle = \langle X^2 \rangle - \langle X \rangle^2 = 0$$

o, alternativamente, $\delta = 0$. Una distribuzione piú generale soddisfacente alla richiesta che δ sia piccolo è disegnata nella figura 4.1*a*. La distribuzione poissoniana incontrata nel capitolo II, § 8, appartiene infatti a questa classe. Inoltre, il calcolo esplicito di δ sviluppato alla fine del paragrafo citato mostra che per questa distribuzione $\delta = 1/N^{1/2}$, dove N è un numero abbastanza grande che misura la grandezza del sistema (attraverso il numero medio di particelle contenute in un volume V o attraverso il valore dello stesso V). Ora, in un sistema di interesse pratico la dimensione lineare di V è molto grande rispetto alla distanza media tra le molecole. Formalmente, questo viene espresso dal *limite termodinamico*, $V \to \infty$. Si vede in questo caso che δ tende a zero. In altre parole, esiste una separazione netta tra comportamento macroscopico e fluttuazioni. Molte delle altre distribuzioni fondamentali incontrate nella teoria della probabilità, come, ad esempio, le distribuzioni binomiale e gaussiana, condividono questa importante proprietà.

I precedenti aspetti possono essere espressi in una forma elegante e generale attraverso due risultati fondamentali della teoria della probabilità noti come la *legge dei grandi numeri* e il *teorema del limite centrale*. In entrambi i teoremi si è interessati alle proprietà della variabile X che è la somma delle variabili casuali, statisticamente indipendenti con una distribuzione comune:

$$X = X_1 + X_2 + \cdots + X_n \tag{4.5}$$

Sotto l'ipotesi molto debole che il valor medio, $\langle X_k \rangle \equiv m$, delle singole distribuzioni esista, la legge dei grandi numeri afferma che la probabilità che la media aritmetica, $(X_1 + \cdots + X_n)/n$, differisca dalla media statistica, m, per meno di un prescritto valore ε, tende a uno quando

$n \to \infty$ e $\varepsilon \to 0$. Sotto l'ulteriore ipotesi che la varianza individuale $\langle \delta X_n^2 \rangle \equiv \sigma^2$ esista, il teorema del limite centrale afferma che nel limite $n \to \infty$ la distribuzione di probabilità di $[(X_1 + \cdots + X_n)/n - m]/(\sigma\sqrt{n})$ si avvicina arbitrariamente a una *distribuzione normale*, l'integrale della *funzione gaussiana di densità*. Inoltre la varianza di $(X_1 + \cdots + X_n)/n$ è dell'ordine di $\sigma^2 n^{-1}$, mentre la varianza di X stessa è $\sigma^2 n$. Questa è esattamente la stima alla quale siamo giunti attraverso il calcolo esplicito della varianza poissoniana, dove il ruolo di n viene giocato dal parametro di grandezza N esprimendo il fatto che la variabile X è la somma dei valori parziali degli osservabili relativi a piccoli elementi di volume del sistema. Questa X è detta *variabile estensiva*.

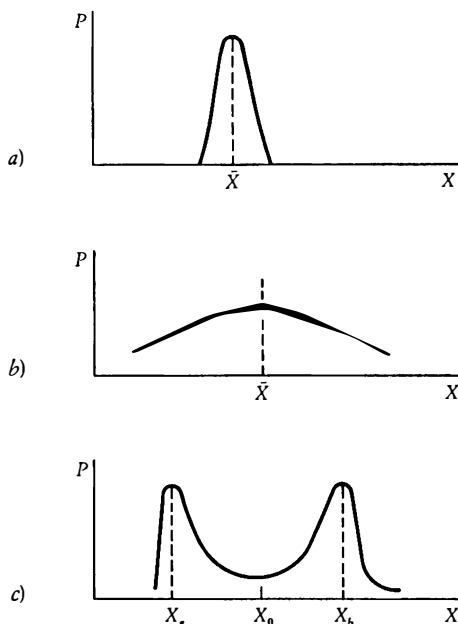
Risultati di questo tipo dominano ancora la letteratura in teoria del-

Figura 4.1

Analogo stocastico della biforcazione. Non appena il sistema incrocia il valore critico λ_c di un parametro di controllo λ, la funzione di probabilità cambia forma.

a) $\lambda < \lambda_c$: una forma unimodale fortemente piccata in corrispondenza di un unico attrattore.

b) $\lambda = \lambda_c$: una distribuzione piatta.

c) $\lambda > \lambda_c$: una distribuzione con molte gobbe i cui massimi coincidono con i nuovi attrattori che emergono dopo la biforcazione.

la probabilità. È comunque molto facile costruire controesempi. Si consideri, per esempio, una distribuzione di probabilità di una macrovariabile (una variabile estensiva, o una variabile intensiva costruita dividendo una variabile estensiva per la dimensione del sistema) come mostrato in figura 4.1*c*. Chiaramente questa non è una gaussiana. È una distribuzione a due gobbe, i cui due massimi per $X = X_a$ e $X = X_b$, sono separati da un minimo in $X = X_0$. Idealizziamo ulteriormente la situazione assimilando P a due picchi infinitamente acuti.

(4.6) $$P = \frac{1}{2}(\delta^{kr}_{X,X_a} + \delta^{kr}_{X,X_b})$$

dove il fattore $\frac{1}{2}$ assicura che $P(X)$ è normalizzata a uno,

$$\sum_{X=0}^{\infty} P(X) = 1$$

Calcoliamo ora il valor medio e la varianza. Otteniamo cosí:

$$\langle X \rangle = \sum_X X\,P(X) = \frac{1}{2}(X_a + X_b)$$

$$\langle X^2 \rangle = \sum_X X^2 P(X) = \frac{1}{2}(X_a^2 + X_b^2)$$

e

$$\langle \delta X^2 \rangle = \frac{1}{2}(X_a^2 + X_b^2) - \frac{1}{4}(X_a + X_b)^2$$

$$= \frac{1}{4}(X_a - X_b)^2$$

Questa conduce a una dispersione statistica

$$\delta = \left(\frac{\langle \delta X^2 \rangle}{\langle X \rangle^2}\right)^{1/2} = \frac{(X_b - X_a)}{(X_b + X_a)}$$

In un sistema macroscopico la variabile X è *estensiva*, che equivale a dire, proporzionale alla grandezza del sistema N. Esprimiamo ciò ponendo

(4.7) $$X = N \cdot x$$

dove la *variabile intensiva* x rimane finita. Questo porta alla seguente stima di δ:

$$\delta = \frac{(x_b - x_a)}{(x_b + x_a)} = \text{quantità finita di ordine uno.} \tag{4.8}$$

In altri termini la varianza è dello stesso ordine della media, e conseguentemente l'effetto delle fluttuazioni raggiunge un livello macroscopico. Stiamo violando chiaramente le premesse del teorema del limite centrale, e conseguentemente le variabili distribuite secondo la figura 4.1*c* non possono in nessun modo essere viste come la somma di variabili casuali statisticamente indipendenti. Possiamo già sospettare che questo debba riflettere l'apparire di *comportamenti coerenti* all'interno del sistema. Un esame alle figure 4.1*a* e 4.1*c* suggerisce che questa coerenza debba essere attribuita a una transizione da uno stato caratterizzato da un unico valore piú probabile $\bar{X}$ a uno stato caratterizzato da due tali valori, X_a e X_b. Questo non è altro che l'*analogo stocastico della biforcazione*, il fenomeno che è stato al centro delle nostre preoccupazioni nel capitolo III. Chiaramente nel contesto di una descrizione deterministica il problema di come questa coerenza emerga e si mantenga non può nemmeno essere formulato. Questa è una delle ragioni base per cui lo studio delle fluttuazioni costituisce una parte integrale della dinamica dei fenomeni complessi. Un'altra ragione sta nel fatto che, in un regime a molte gobbe oltre la biforcazione, è importante stimare la stabilità relativa degli stati piú probabili, e il tempo che il sistema passa nelle loro vicinanze prima di eseguire una transizione in un altro dei rami simultaneamente stabili. Nuovamente, questo problema può essere formulato solo nel contesto di una descrizione allargata che incorpori l'effetto delle fluttuazioni.

La figura 4.1*b* descrive la forma della distribuzione di probabilità che ci si aspetta nel limite tra regimi unimodali e multimodali. Si ha una distribuzione piatta, che «prepara la strada» della biforcazione esplorando con una apprezzabile probabilità regioni dello spazio delle fasi che sono abbastanza lontane dal valore piú probabile. Questo *analogo probabilistico della stabilità marginale* sarà ulteriormente analizzato nel § 4 dove si mostrerà anche che porta al crollo del teorema del limite centrale. Fino ad ora in questo paragrafo abbiamo focalizzato la nostra discussione sul ruolo delle fluttuazioni generate spontaneamente dalla dinamica molecolare del sistema. Abbiamo notato comunque nel capitolo III, §§ 7-10, che in vaste classi di sistemi dinamici il comportamento ca-

suale può anche sorgere attraverso la biforcazione di moti instabili che conduce, per esempio, ad attrattori caotici. Considerazioni probabilistiche diventano quindi ancora un supplemento indispensabile dell'approccio deterministico che permette di determinare, per esempio, la frequenza con la quale vengono incontrate differenti parti dell'attrattore caotico, o il tempo oltre il quale svanirà la memoria di un particolare stato iniziale.

La principale differenza tra le fluttuazioni e gli avvenimenti casuali associati al caos sta nella *forza* e nella *scala temporale* delle deviazioni dalla media. Come abbiamo visto precedentemente in questo paragrafo, quando un sistema fisico-chimico ha un unico attrattore asintoticamente stabile le fluttuazioni attorno al valor medio sono estremamente piccole, all'incirca $1/N^{1/2}$ volte meno dei valori deterministici. Persino in un regime multimodale nel quale la dispersione attorno alla media è grande, la dispersione attorno a ogni picco locale rimane molto piccola. Inoltre, i valori delle probabilità nella vicinanza dello stato instabile sono piccoli come e^{-N}, l'esponenziale inverso del parametro di grandezza N. Questo significa che per la maggior parte del tempo il sistema sta nelle immediate vicinanze di un picco di probabilità e che il *tempo di transizione* tra i picchi è molto lungo, dell'ordine di e^{N}. Nessuna di queste proprietà vale per un tipico attrattore caotico. Qui il sistema passa attraverso stati distanti durante un lasso di tempo comparabile alle scale dei tempi deterministici, questo a indicare che la forza delle deviazioni dalla media è paragonabile alla media stessa. Questo ci ricorda certamente il comportamento associato a un moto turbolento e sottolinea una volta ancora il ruolo del caos come un modello naturale di turbolenza nei sistemi fisico-chimici.

Una proprietà simile caratterizza la dinamica dei sistemi soggetti a condizioni ambientali casuali, la cui variabilità è percepita dal sistema come *un rumore esterno*. Rimandiamo il lettore al libro di W. Horsthemke e R. Lefever, *Noise-induced Transitions* (Berlin 1984) per un resoconto dettagliato su questo importante argomento.

2. *Processi di Markov. Equazione principale.*

Nel paragrafo precedente abbiamo ragionato come se la distribuzione di probabilità che descrive il sistema a cui siamo interessati fosse nota, almeno per quanto concerne i suoi aspetti generali. In alcuni casi questo è ciò che infatti avviene. Per esempio l'analisi nel capitolo II,

§ 8, mostrava che in un sistema ideale in equilibrio termodinamico, la probabilità di trovare un dato numero di particelle in un piccolo volume è poissoniana. Analogamente, se in un certo sistema la validità delle condizioni per il teorema del limite centrale, menzionate nel § 1, è garantita, allora si sa che le corrispondenti variabili saranno descritte dalla distribuzione normale. In generale comunque tutto ciò che è dato sono i processi che si stanno svolgendo in un sistema dinamico: conduzione e convezione termica nel problema di Bénard, reazioni e diffusione in un sistema chimico isotermo, e cosí via. Come si può determinare, sulla base di queste informazioni, la forma della sottostante distribuzione di probabilità? Per rispondere a questa domanda, è conveniente costruire un'equazione del bilancio che conta i processi che conducono il sistema a un certo stato Q, e i processi che lo rimuovono da questo stato. Ovviamente,

$$(4.9)\quad \frac{d}{dt}\text{Prob}\,(Q, t) = (\text{numero di transizioni verso lo stato } Q) - (\text{numero delle transizioni da } Q \text{ per unità di tempo}) \equiv R_+(Q) - R_-(Q)$$

e il problema è ora ridotto alla determinazione delle quantità di transizione R_+ e R_-. Qualunque sia la forma esplicita di queste quantità essa dovrà essere compatibile con certi vincoli. Per esempio, è noto che i sistemi fisico-chimici in equilibrio termodinamico soddisfano la condizione del *bilancio dettagliato*, alla quale abbiamo ripetutamente alluso nei capitoli I e II. Dunque, se si decompongono R_+ e R_- nei processi elementari che avvengono nel sistema,

$$R_\pm = \sum_k r_{k,\pm}$$

deve essere soddisfatta la seguente condizione:

$$(4.10)\qquad (r_{k,+})_{equil} = (r_{k,-})_{equil}$$

Queste relazioni devono essere a loro volta compatibili con la forma della distribuzione di probabilità nello stato di equilibrio, che è nota dalla meccanica statistica. Un caso limite di tale distribuzione è la poissoniana. Piú generalmente, come mostrato inizialmente da Einstein, all'equilibrio la probabilità di una fluttuazione è interamente determina-

ta da quantità termodinamiche. In un sistema isolato l'inversa della famosa formula di Boltzmann,

$$S = k_B \ln (\text{numero di disposizioni molecolare compatibili con un dato valore dell'energia})$$

porta a

$$P_{eq} \sim \exp \frac{1}{k_B} \Delta S \tag{4.11}$$

dove ΔS è la variazione di entropia dovuta alla fluttuazione,

$$\Delta S = S(Q) - S(Q_{equil}).$$

Un altro vincolo che l'equazione 4.9 dovrebbe soddisfare è di ridursi, in un certo senso restrittivo, alle equazioni di evoluzione che intervengono nella descrizione deterministica, come le equazioni della fluidomeccanica e della cinetica chimica. Ci si aspetta che un'osservazione macroscopica di un sistema fisico-chimico offrirà valori rappresentativi del suo stato piú probabile. Dobbiamo richiedere dunque che i picchi di $P(Q, t)$ siano soluzioni delle equazioni deterministiche. Se la distribuzione è unimodale (cfr. figura 4.1*a*), questo implica a sua volta che l'equazione del valor medio è vicina all'equazione deterministica e che la correzione è essenzialmente proporzionale a una potenza inversa della grandezza del sistema.

Nonostante il loro interesse, comunque, i precedenti vincoli non sono sufficienti a fissare in modo unico la forma dell'equazione 4.9. Dovremo pertanto appellarci ora a qualche ragione intuitiva. Abbiamo già sottolineato che le fluttuazioni nascono dal moto casuale delle molecole, il quale è dovuto alle interazioni intermolecolari e al grande numero di particelle coinvolte nella dinamica dei principali sistemi fisico-chimici di interesse. Esse sono quindi eventi essenzialmente localizzati il cui spazio caratteristico e scala temporale sono estremamente piccoli. In queste circostanze, ci si aspetta che le quantità di transizione che appaiono nell'equazione 4.9 dipenderanno solo dallo stato Q e dagli stati che possono essere connessi a Q attraverso un'unica fluttuazione. In altri termini, supponiamo che la memoria di transizioni consecutive che nascono da fluttuazioni consecutive, venga perduta dopo la prima transizione. Questa condizione definisce una classe estremamente importante di fenomeni noti come *processi di Markov* il cui studio è uno dei principali problemi della teoria della probabilità.

Una semplicissima illustrazione dei processi di Markov viene mo-

strata nella figura 4.2. Il sistema considerato possiede 5 stati, da I a V. I e V sono «frontiere assorbenti», nel senso che il sistema vi rimane una volta che le ha raggiunte. D'altra parte, una volta giunto nello stato II, III, IV, il sistema balza immediatamente a destra o a sinistra con probabilità q e $1-q$ rispettivamente. Nel limite in cui il numero degli stati è molto grande, questo «cammino casuale» diviene una rappresentazione realistica del moto di una particella in un fluido ospite, e conduce alla familiare legge di Fick della diffusione.

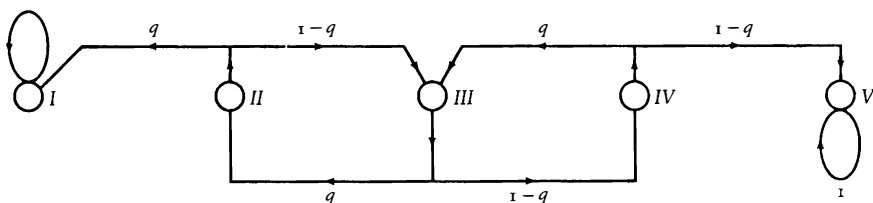
Nei sistemi che presentano un comportamento complesso, si ha naturalmente a che fare con situazioni piú intricate, a causa della nonlinearità della cinetica. Ciò nonostante, si può scrivere una forma esplicita dell'equazione 4.9, nota come *equazione principale* (*master equation*). Essa esprime l'indice $R_+(Q)$ come il prodotto della probabilità di transizione, per unità di tempo, per passare dallo stato Q' a Q, per la probabilità di essere nello stato Q' al tempo t in primo luogo, sommato su tutti gli stati Q' che possono condurre a Q in un unico passo, in virtú di processi dinamici elementari che avvengono nel sistema. Analogamente, $R_-(Q)$ è il prodotto della probabilità di essere nello stato Q al tempo t in primo luogo per la somma delle probabilità di transizione per unità di tempo da Q a tutti gli stati Q' accessibili da Q.

Scriviamo questo bilancio nella forma:

$$\frac{dP(Q,t)}{dt} = \sum_{Q' \neq Q} [W(Q|Q')P(Q',t) - W(Q'|Q)P(Q,t)] \tag{4.12a}$$

dove la probabilità di transizione per unità di tempo $W(Q|Q')$ è non negativa per ogni $Q' \neq Q$. Inoltre, poiché $P(Q, t)$ deve essere normaliz-

Figura 4.2

Illustrazione di un processo di Markov. Le frecce fuoriuscenti dagli stati da I a V indicano la direzione nella quale avvengono le transizioni permesse.

zato a uno per tutti i valori del tempo, W deve soddisfare la condizione

$$\sum_{Q} W(Q|Q') = 0 \qquad (4.12b)$$

che esprime il fatto intuitivamente chiaro che, partendo dallo stato Q', tutto ciò che il sistema può fare è o stare in Q' o compiere una transizione agli stati accessibili $Q \neq Q'$.

L'equazione 4.12*a* è lineare rispetto alla funzione incognita $P(Q, t)$, ma questo non implica che la sua soluzione sia facile. Infatti, la complessità delle leggi della dinamica dell'evoluzione (ad esempio i *feedback* presenti in un sistema chimico e biologico), è riflessa dalla presenza di non-linearità nei tassi di transizione W. Vicino all'equilibrio questi *feedback* sono resi inefficaci dal bilancio dettagliato, ma lontano dall'equilibrio essi si manifestano pienamente. Ne consegue che l'analisi completa dell'equazione 4.12 costituisce un problema aperto. Inoltre, eseguendo calcoli perturbativi, cercando i semplici casi limite esattamente risolubili, ed eseguendo simulazioni al calcolatore, si giunge alla descrizione generale delle leggi che governano le fluttuazioni nei sistemi non-lineari di non-equilibrio. I risultati principali verranno presentati nei successivi paragrafi di questo capitolo. Chiuderemo questo paragrafo con alcuni esempi concreti della struttura dell'equazione principale.

Una classe di fenomeni riconducibili alla equazione 4.12 è fornita dai *processi di natalità e mortalità*. Questi sono fenomeni nei quali lo stato Q è interamente caratterizzato da un insieme di interi $\{X_\alpha\}$, che possono unicamente variare da 1 a -1. Esempi ovvi di tale fenomeno sono le reazioni chimiche e biochimiche, nelle quali $\{X_\alpha\}$ rappresenta il numero di molecole della specie a presenti. Considerando per semplicità una singola specie, possiamo scrivere

$$W(X|X') = \sum_{\rho} W_\rho(X|X - r_\rho)\delta^{Kr}_{X', X-r_\rho} \qquad r_\rho = 0, \pm 1 \qquad (4.13a)$$

dove ρ indica un particolare passo della reazione e r_ρ rappresenta il numero di molecole consumate ($r_\rho < 0$) oppure create ($r_\rho > 0$) nel processo. Definiamo gli indici di natalità e di mortalità, rispettivamente λ e μ, nel modo seguente:

$$\begin{aligned} \lambda_X &= \sum_{\rho} W_\rho(X|X - 1) \equiv N\lambda(X/N) \qquad X \geqslant 0 \\ \mu_X &= \sum_{\rho} W_\rho(X|X + 1) \equiv N\mu(X/N) \qquad X \geqslant 1 \end{aligned} \qquad (4.13b)$$

L'ultima parte delle precedenti uguaglianze enfatizza un importante aspetto della dinamica stocastica: in seguito al carattere a corto raggio delle interazioni, il numero di transizioni per unità di tempo che avvengono in un sistema di grandezza N deve essere proporzionale alla grandezza stessa. Ovvero, ogni piccolo elemento del sistema può solo percepire il suo immediato intorno; dunque, sommando sopra tutti gli elementi, si otterrà un fattore N che moltiplica una funzione della variabile intensiva X/N. Naturalmente assumiamo che stiamo trattando interazioni di origine fisico-chimica. Nella dinamica dei sistemi sociali la situazione può essere diversa. A causa delle telecomunicazioni, dei mass-media, eccetera, ogni elemento del sistema può essere sensibile a tutti gli altri elementi presenti. Questo condurrebbe a indici di transizione proporzionali a opportune potenze N^a, dove l'esponente a può essere maggiore di uno.

Illustreremo ora la struttura delle probabilità di transizione per unità di tempo nel modello autocatalitico dell'equazione 2.15. Dobbiamo calcolare le quattro quantità W_1, W_{-1}, W_2, W_{-2}, che descrivono successivamente i tassi di avanzamento e di arretramento delle due reazioni chimiche coinvolte nel modello. Consideriamo dapprima W_1. La reazione $A + 2X \xrightarrow{k_1} 3X$ nasce dall'incontro di coppie di particelle X con particelle di specie A. Conseguentemente,

$$W_1(X + 1|X) = k_1 \frac{AX(X-1)}{N^2} \qquad (\text{passo } A + 2X \xrightarrow{k_1} 3X)$$

dove il fattore $1/N^2$ assicura che l'intera espressione è estensiva. Analogamente,

$$W_{-1}(X - 1|X) = k_2 \frac{X(X-1)(X-2)}{N^2} \qquad (\text{passo } 3X \xrightarrow{k_2} A + 2X)$$

$$W_2(X - 1|X) = k_3 X \qquad (\text{passo } X \xrightarrow{k_3} B)$$

$$W_{-2}(X + 1|X) = k_4 B \qquad (\text{passo } B \xrightarrow{k_4} X)$$

Gli indici di natalità e di mortalità sono dati da

$$\lambda_X = W_1 + W_{-2}$$
$$\mu_X = W_{-1} + W_2$$

Riassumendo, dalle equazioni 4.13 e 4.12 otteniamo la seguente forma dell'equazione principale per un processo di natalità e mortalità:

$$\frac{dP(X,t)}{dt} = \lambda_{X-1}P(X-1,t) + \mu_{X+1}P(X+1,t) - (\lambda_X + \mu_X)P(X,t) \tag{4.14}$$

È istruttivo verificare, con questa forma abbastanza esplicita, la connessione tra le descrizioni stocastiche e quelle macroscopiche. Moltiplicando entrambi i membri per X e sommando poi rispetto a X, si ottiene dalla descrizione stocastica

$$\frac{d\langle X\rangle}{dt} = \langle\lambda_X\rangle - \langle\mu_X\rangle$$

mentre la descrizione macroscopica dello stesso sistema ci dà:

$$\frac{d\bar{X}}{dt} = \text{parte dominante di } (\lambda_{\bar{X}} - \mu_{\bar{X}}) \text{ nel limite di grandezza } \bar{X}$$

Le due descrizioni dunque differiscono se la media di λ o di μ è diversa dal valore di queste funzioni calcolate per $X = \langle X\rangle$. Questo è precisamente quello che succede in un sistema non-lineare. Per vedere la natura della correzione alla descrizione macroscopica, consideriamo il semplice esempio di un processo di pura mortalità,

$$2X \xrightarrow{k} X + E$$

L'indice di mortalità μ_X conta la frequenza degli incontri di coppie di particelle X. È dunque dato da $\mu_X = k(1/N)X(X-1)$. Conseguentemente si ha

$$\begin{aligned}\frac{d\langle X\rangle}{dt} &= -k\frac{1}{N}\langle X(X-1)\rangle \\ &= -k\frac{1}{N}(\langle X\rangle^2 + \langle\delta X^2\rangle - \langle X\rangle)\end{aligned}$$

Passando alla variabile intensiva, $x = X/N$, possiamo scrivere la precedente relazione come

$$\frac{d\langle x\rangle}{dt} = \underbrace{-k\langle x\rangle^2}_{\text{Legge macroscopica}} \quad \underbrace{-\frac{k}{N}\frac{\langle\delta X^2\rangle - \langle X\rangle}{N}}_{\text{Correzione dovuta alle fluttuazioni}} \tag{4.15}$$

Vediamo anche che la deviazione dalla legge macroscopica è proporzionale alla deviazione della sottostante distribuzione di probabilità dalla poissoniana, poiché in quest'ultima distribuzione $\langle \delta X^2 \rangle = \langle X \rangle$ (cfr. equazione 2.28). Questa conclusione è di fatto abbastanza generale ed essenzialmente indipendente dai dettagli della cinetica. Tornando al nostro esempio, in un sistema descritto da una distribuzione unimodale $\langle \delta X^2 \rangle$ è dell'ordine di N, per il teorema del limite centrale. La correzione alla legge macroscopica è allora pesata dall'inverso del parametro di grandezza N ed è piccola se N è grande. Quest'ultimo tuttavia non è necessariamente il nostro caso. In contrasto con le assunzioni semplificatrici che portano alla forma dell'equazione 4.13 per le probabilità di transizione, la maggior parte delle fluttuazioni è data in realtà da eventi localizzati; pertanto N sarà tipicamente la grandezza di una regione la cui dimensione caratteristica è dello stesso ordine delle fluttuazioni. Questo argomento mostra che l'effetto delle fluttuazioni non è universale, ma dipende in modo cruciale dal loro ordine di grandezza.

La necessità di un *approccio locale* alle fluttuazioni va di fatto ancora piú in là. Per assicurare validità alla descrizione di Markov, è essenziale scegliere le variabili di stato in modo tale che ogni transizione individuale tra gli stati possa essere attribuita direttamente a un'unica fluttuazione «elementare». In una descrizione globale questo non sarà necessariamente cosí, e il sottostante processo stocastico potrà diventare non di Markov. Di fatto nell'equazione 4.14, si dovrebbe intendere che le reazioni chimiche hanno luogo in ogni piccolo elemento di volume ΔV del nostro sistema (cfr. figura 4.3), e che i loro effetti vengono sommati successivamente sopra l'intero volume V. I vari elementi di volume saranno naturalmente statisticamente accoppiati attraverso il trasferimento di particelle, o del momento, o dell'energia attraverso le loro frontiere. Pertanto, l'equazione 4.14 dovrà essere completata da un termine addizionale che descrive questo processo di trasferimento.

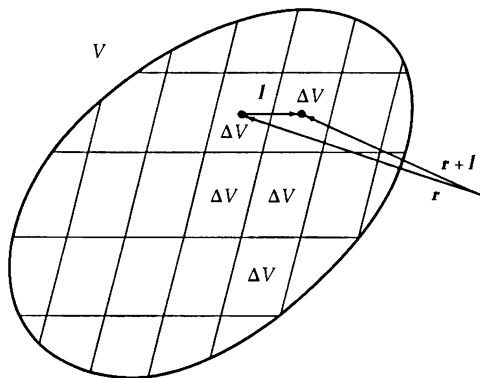
Abbiamo discusso fino ad ora in termini di variabili discrete e di spazi delle fasi discreti. In un sistema fisico questo è infatti il livello fondamentale di descrizione, poiché, dopotutto, le fluttuazioni nascono a causa del carattere discreto della materia. In molti casi, comunque, il sottostante processo stocastico discreto si può ridurre, in un certo ben definito senso asintotico, a un processo stocastico con realizzazioni continue. In questo limite la descrizione diviene molto simile a quella macroscopica, e la differenza risiede nel fatto che il tasso macroscopico è ora completato da una *forza stocastica* che esprime l'effetto delle fluttuazioni. Questa descrizione, che risale al classico studio di Paul Lange-

vin sui moti browniani, si dimostra essere in molte occasioni un fertile approccio all'analisi stocastica di vaste classi di sistemi dinamici.

3. *Processi di Markov e irreversibilità.*

È probabile che il lettore sia stato colpito dall'apparente somiglianza tra i primi due paragrafi di questo capitolo e il § 3 del capitolo II. In entrambi i casi si è infatti adottato un linguaggio probabilistico e l'evoluzione della sottostante distribuzione di probabilità è ridotta alla soluzione di una equazione lineare del primo ordine rispetto alla variabile tempo: l'equazione principale 4.12, oppure l'equazione di Liouville 3.7. Nonostante queste analogie tra le due descrizioni, mostreremo ora che i processi di Markov si situano in una posizione completamente diversa. Questo si poteva già intuire dalla relazione stabilita nel precedente paragrafo tra l'equazione principale e le leggi della produzione macroscopica dei sistemi dissipativi, una proprietà che è da contrapporre all'invarianza per inversione temporale dell'equazione di Liouville. Come vedremo ora, la connessione tra i processi di Markov e i sistemi

Figura 4.3

Partizione di un sistema macroscopico in sottovolumi di grandezza ΔV centrati nei punti $\boldsymbol{r}$ dello spazio accoppiati ai loro vicini $\boldsymbol{r} + \boldsymbol{l}$ attraverso il trasferimento di massa, di momento o di energia. A causa del carattere locale delle fluttuazioni, ci si aspetta che, scegliendo in modo opportuno ΔV, il procedimento consistente nel seguire l'evoluzione delle variabili di stato all'interno dell'insieme di tutti i ΔV sia markoviano.

dissipativi è infatti un aspetto molto generale, profondamente radicato nella struttura stessa dell'equazione principale. Piú specificatamente mostreremo che, se il tempo cresce all'infinito, le soluzioni dell'equazione principale mostrano una universale tendenza verso un unico stato stazionario finale, condizionato da una *funzione di Ljapunov* molto simile alle funzioni termodinamiche considerate nel capitolo II, § 6.

Cominciamo con l'introdurre l'*entropia* di un processo di Markov. La necessità di tale quantità apparve abbastanza presto nella teoria della probabilità. Fu essenzialmente motivata dallo sviluppo di modelli probabilistici esattamente risolubili che si proponevano di riconciliare l'andamento irreversibile verso l'equilibrio predetto dalla teoria cinetica di Boltzmann e la reversibilità delle leggi del moto dei sistemi conservativi. Fin dall'avvento della teoria dell'informazione, l'entropia ha dimostrato di essere anche un importante concetto nella teoria matematica della comunicazione. Discuteremo questo aspetto dell'entropia piú avanti (§ 7). Qui osserviamo solamente che, in entrambi i contesti fisico-chimico e di teoria dell'informazione, è usuale cercare una quantità che soddisfi un numero di richieste imposte dall'intuizione e dall'analogia con la termodinamica, come ad esempio

1. Per un dato numero di stati, N, e per

$$P(Q) \geqslant 0, \ \sum_Q P(Q) = 1$$

la funzione entropia S assume il suo massimo valore in $P_Q = 1/N$.

2. L'entropia di un sistema composto AB è uguale all'entropia del sottosistema A piú l'entropia (condizionale) $S_A(B)$ del sottosistema B, nella condizione che il sottosistema A sia in un dato stato A:

$$S(AB) = S(A) + S_A(B)$$

Questa proprietà viene solitamente chiamata *subadditività*.

3. L'aggiunta di un evento impossibile, Q_α con $[P(Q_\alpha) = 0]$, non cambia l'entropia del processo.

In queste condizioni, e supponendo inoltre che si richieda anche che S sia una funzione continua delle P, si può stabilire che S è univocamente determinata a meno di una costante moltiplicativa:

(4.16) $$S[P(Q), t] = -\sum_{Q} P(Q, t) \ln P(Q, t)$$

Cosí formulata, questa espressione non è una adeguata funzione di Ljapunov nel caso piú generale di un processo di Markov che descrive un sistema soggetto a vincoli di non-equilibrio. Per giungere a una tale funzione, dobbiamo introdurre una ipotesi addizionale, precisamente che il presente processo di Markov possieda una distribuzione stazionaria di probabilità, $P_S(Q)$, che soddisfi identicamente l'equazione principale 4.12*a* e con $dP_S/dt = 0$. Non tutti i processi di Markov sono compatibili con questi requisiti: è infatti noto dalla teoria della probabilità che questa proprietà è un aspetto rilevante dei processi stocastici i cui stati sono *ergodici*; in altri termini, tutti gli stati sono accessibili con probabilità uno, e il tempo medio di ritorno a uno stato particolare partendo da esso (noto come il *tempo di ricorrenza*) è finito.

Con l'aiuto della funzione $PS(Q)$, si può ora costruire la quantità

(4.17) $$K(P(Q), t) = \sum_{Q} P(Q, t) \ln \frac{P(Q, t)}{P_S(Q)}$$

che si annulla ovviamente per $P = P_S$ e rimane positiva altrimenti. La sua variazione temporale può essere calcolata per mezzo dell'equazione 4.12*a*:

$$\frac{dK}{dt} = \sum_{\substack{Q, Q' \\ (Q \neq Q')}} [W(Q|Q')P(Q', t) - W(Q'|Q)P(Q, t)] \ln \frac{P(Q, t)}{P_S(Q)}$$

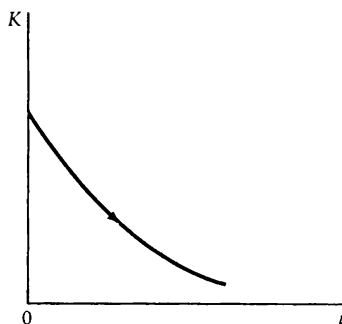
Permutando gli indici Q e Q', e usando la positività di $W(Q|Q')$ per $Q \neq Q'$ cosí come la ben nota proprietà del logaritmo, $\ln x \leqslant x - 1$ per ogni x reale, si ottiene finalmente

(4.18) $$\frac{dK}{dt} \leqslant 0$$

Questo risultato descritto nella figura 4.4, produce una sorprendente somiglianza con i risultati del capitolo II, §§ 5 e 6 (si veda specialmente la figura 2.7), concernenti il comportamento monotòno dell'entropia in un sistema isolato imposto dalla seconda legge della termodinamica. Si dovrebbe notare che K *non* è identico all'entropia. Inoltre la disuguaglianza 4.18 si applica altrettanto bene nei sistemi non isolati soggetti a vincoli permanenti di non-equilibrio. Tuttavia, cosí come la seconda

legge in un sistema isolato esprime l'universale tendenza a un unico stato di equilibrio, anche la disuguaglianza 4.18 esprime che P_S è unico e globalmente stabile, nel senso che ogni condizione iniziale $P(Q, t)$ è obbligata a evolvere verso di esso quando $t \rightarrow \infty$. In altre parole, il problema dell'instabilità e della molteplicità di soluzioni familiare dallo studio delle biforcazioni in sistemi dinamici non-lineari semplicemente non sorge a livello probabilistico. Un'occhiata alla figura 4.1 ci mostra perché deve essere cosí: la biforcazione viene riflessa da un cambiamento qualitativo nella struttura della sottostante distribuzione di probabilità, come, per esempio, l'apparire di gobbe multiple, piuttosto che dalla molteplicità della stessa distribuzione di probabilità. Una notevole eccezione comunque, nasce nel limite per il quale la grandezza del sistema tende all'infinito. La larghezza di ogni picco di probabilità diviene allora zero, e la distribuzione di probabilità si decompone in un numero di distribuzioni infinitamente strette (un esempio di questo è già stato fornito dall'equazione 4.6). I vari stati del processo di Markov non definiscono piú un insieme ergodico, e differenti condizioni iniziali sono attratte verso differenti distribuzioni di probabilità finali. Si può dimostrare che questa situazione descrive sia le transizioni di equilibrio che di non-equilibrio in sistemi infinitamente estesi.

Un'altra conseguenza molto importante della disuguaglianza 4.18 è di mostrare come l'equazione principale, equazione 4.12*a*, equivalga a una *contrazione* nello spazio delle distribuzioni di probabilità, poiché, qualunque siano i dettagli della dinamica, la loro «distanza» da $P_S(Q)$ è

Figura 4.4
Irreversibilità di un processo di Markov descritto dall'equazione principale (cfr. equazione 4.12). La funzione eccesso di entropia, K (cfr. equazione 4.17), decresce monotonamente nel tempo.

obbligata a decrescere. Questa proprietà, assieme alla linearità della equazione principale è alla base della profonda connessione fra i processi di Markov e l'importante concetto matematico di *semigruppi dissipativi.*

Per capire intuitivamente questa connessione, paragoniamo le seguenti due equazioni di evoluzione:

(4.19*a*) $$\frac{dy}{dt} = -\lambda y \qquad (t>0; \lambda>0)$$

(4.19*b*) $$\frac{d^2x}{dt^2} = -\omega_0^2 x$$

Queste equazioni descrivono rispettivamente un decadimento irreversibile verso uno stato finale asintoticamente stabile, $y = 0$, in un sistema dissipativo (si veda per esempio il capitolo II, § 2), e una tipica evoluzione reversibile nel tempo come un oscillatore armonico (cfr. capitolo III, § 3). Possiamo anche scrivere, invece della seconda equazione 4.19, la corrispondente equazione di Liouville nelle variabili d'azione e d'angolo (cfr. equazione 3.7 e capitolo III, § 3):

(4.20) $$\frac{\partial \rho}{\partial t} = -\omega \frac{\partial \rho}{\partial \varphi}$$

Le soluzioni delle equazioni apparentemente simili 4.19*a* e 4.20 sono in effetti completamente differenti. Per la 4.19*a* si ha semplicemente

(4.21*a*) $$y(t) = e^{-\lambda t} y(0)$$

mentre per la 4.20 si deve costruire una soluzione compatibile con il fatto che φ rappresenti un angolo. Per fare ciò è necessario introdurre l'unità immaginaria, $i = \sqrt{-1}$:

(4.21*b*) $$\rho(t) = e^{i(\omega t - \varphi)} \rho(0)$$

che rimane infatti invariante quando φ varia di una quantità pari a $\pm 2\pi$, $\pm 4\pi$, ...

Entrambe le equazioni. 4.21*a* e 4.21*b* mostrano un «operatore d'evoluzione» che agisce sulla condizione iniziale per produrre il valore istantaneo, che è della forma

(4.22*a*) $$U_t = \exp(tL)$$

e soddisfa la cosiddetta proprietà di gruppo

$$(4.22b) \qquad U_t U_{t'} = \exp[(t + t')L] = U_{t+t'}$$

Tuttavia, nella 4.21*a* $L = -\lambda$ è un numero reale negativo. Questo implica che U_t ha un buon comportamento solamente per $t>0$, nel qual caso, in virtú della 4.21*a*, otteniamo un decadimento monotòno delle variabili di stato verso un unico stato stazionario. In termini piú astratti, diremo che il corrispondente operatore U_t definisce un semigruppo dissipativo. D'altra parte, il valore assoluto di ρ (la sua «misura « in termini piú astratti) nell'equazione 4.21*b* rimane immutata: $|\rho(t)| = |\rho(0)|$. Diremo che il corrispondente operatore U_t definisce un *gruppo unitario*. Si noti che in quest'ultimo caso l'invarianza per inversione temporale della sottostante dinamica viene riflessa dalla proprietà $U_t = U^*_{-t'}$; in altre parole, l'inversione temporale a livello di U_t deve essere accompagnata dal coniugio complesso.

Un rapido esame ci convincerà che i due semplici modelli di cui sopra sono concettualmente simili all'equazione principale e all'equazione di Liouville per un sistema conservativo. La sola differenza sta nel fatto che i corrispondenti L nell'esponenziale dell'equazione 4.22*a* non sono piú numeri ma operatori. Possiamo quindi affermare che il *marchio dell'irreversibilità* sta nella comparsa di un semigruppo dissipativo che descrive un processo di Markov appropriatamente definito. Questo a sua volta porta a uno dei piú profondi problemi della fisica, e precisamente: può un tale processo di Markov nascere da una dinamica reversibile nel tempo? Se la risposta a questa domanda risulta essere affermativa, la transizione da un gruppo unitario a un semigruppo dissipativo dovrebbe equivalere a una *rottura della simmetria temporale*: i sistemi conservativi sarebbero quindi gli «antenati» dei sistemi dissipativi, i quali mostrano questo speciale tipo di rottura di simmetria, proprio come lo stato stazionario uniforme di un reagente chimico è l'antenato della struttura dissipativa che descrive una struttura o un ritmo dopo un opportuno punto di biforcazione. Il capitolo v è dedicato alla discussione di questo problema fondamentale.

4. *Correlazioni spaziali e comportamento critico.*

Avendo discusso i fondamentali aspetti della descrizione probabilistica dei sistemi dinamici, possiamo ora rivolgere l'attenzione alle ap-

plicazioni di questi concetti alle varie questioni sollevate nel corso di questa monografia connesse alla comparsa del comportamento complesso.

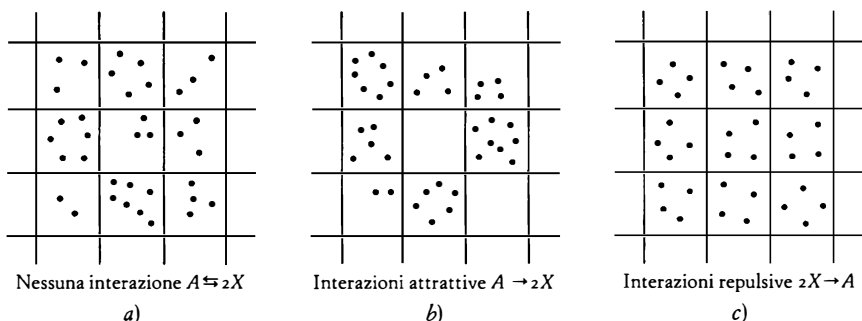
Abbiamo ripetutamente parlato dei fenomeni di transizione lontani dall'equilibrio come associati a degli stati caratterizzati da correlazioni di dimensioni macroscopiche. Nel capitolo II, § 8, abbiamo sottolineato come tali correlazioni possano nascere solamente nei sistemi nei quali il regime delle fluttuazioni poissoniane è stato superato grazie al mantenimento di opportuni vincoli di non-equilibrio. Inoltre nel § 1 di questo capitolo siamo andati piú in là, associando la coerenza implicata dalla biforcazione al venir meno del teorema del limite centrale e al concomitante aumento delle fluttuazioni. In entrambi i casi, deve essere ben chiaro che la coerenza macroscopica e le correlazioni di cui stiamo parlando hanno poco a che fare con le forze di interazione intermolecolari, poiché esse possono nascere anche in una soluzione diluita di molecole incapaci di riconoscersi l'una con l'altra oltre una distanza maggiore di pochi angstrom. In questo paragrafo spieghiamo come può aver luogo questo straordinario fenomeno, usando le proprietà dell'equazione principale e appellandoci continuamente alle analogie e alle differenze con i fenomeni di transizione di fase all'equilibrio.

Come si è rimarcato nel capitolo I, § 9, una transizione di fase all'equilibrio come il raffreddamento o la magnetizzazione spontanea, è il risultato di una competizione fra le forze intermolecolari, che tendono all'ordine del sistema, e il moto termico casuale delle molecole, che ha l'effetto opposto. Ecco perché le transizioni all'equilibrio vengono tipicamente indotte aumentando la pressione (che favorisce l'effetto delle interazioni molecolari), o diminuendo la temperatura (che diminuisce l'effetto del rumore termico). Di fatto, anche lontano dalla transizione di fase, le forze intermolecolari conferiscono a un sistema relativamente denso un certo grado di ordine, il cui raggio è tuttavia corto poiché esso è interamente condizionato dal raggio d'azione delle forze intermolecolari (cfr. figura 1.18). Un'utile visualizzazione di questa situazione è fornita dalla figura 4.5.

Consideriamo un sistema di riferimento all'equilibrio termodinamico in cui l'effetto delle interazioni intermolecolari possa essere trascurato, come, per esempio, un gas perfetto o una soluzione ideale (figura 4.5*a*). In un tale sistema le fluttuazioni in un dato elemento di volume sono poissoniane, una proprietà che riflette il completo disordine che prevale come risultato del moto termico delle particelle costituenti. Supponiamo ora di dare il via alle interazioni attrattive. Le particelle

tenderanno a formare dei gruppi (figura 4.5*b*), e da ciò deriveranno deviazioni maggiori della loro distribuzione spaziale dal valor medio, $\langle \delta X^2 \rangle > \langle X \rangle$. Se, d'altro canto, le interazioni sono repulsive, le particelle saranno distribuite in modo piú efficiente entro il volume (figura 4.5*c*), e ciò diminuirà la deviazione dalla media. In entrambi i casi l'«ordine» spaziale associato con il comportamento non poissoniano rimane a corto raggio, a meno che il sistema non venga portato vicino a un punto critico di transizione di fase: le correlazioni spaziali acquistano una lunga coda, cosa che riflette la capacità di sottostare a un cambiamento che comprende l'intero sistema.

Consideriamo ora un sistema ideale in grado di funzionare sia lontano che vicino all'equilibrio. Per un tale sistema ci si aspetta che i sopra citati effetti del raggruppamento e del comportamento critico non avranno luogo in nessuna circostanza. All'equilibrio ciò si può esplicitamente controllare: la distribuzione di probabilità è poissoniana, e ne risulta che non esiste una correlazione spaziale. Ma cosa accade se il sistema viene portato lontano dall'equilibrio per mezzo di vincoli opportunamente apprestati? Per essere piú specifici, consideriamo la reazione $A \rightleftarrows 2X$ e supponiamo dapprima che A sia in grande eccesso, cosí che si possa trascurare il passo $2X \rightarrow A$. Per evitare l'ac-

Figura 4.5
Illustrazione dell'origine delle correlazioni spaziali a raggio macroscopico in un sistema ideale di non equilibrio. All'equilibrio *a*) il bilancio dettagliato preclude l'accumulazione di X in regioni privilegiate dello spazio. X è perciò distribuito casualmente (statistica poissoniana), come appunto ci si aspetterebbe in un sistema di particelle non interagenti. Il bilancio dettagliato non è valido lontano dall'equilibrio, quindi possono crearsi *b*) o distruggersi *c*) coppie di X in modo correlato. Queste correlazioni saranno mantenute su una distanza dell'ordine del raggio di diffusione. Il sistema, benché ideale, si comporterà quindi come un sistema di particelle interagenti attraverso forze attrattive *b*), o repulsive *c*), di ampio raggio.

cumulo di X in concentrazioni infinite, introduciamo anche una reazione che lo rimuove dal mezzo. Questo ci porta allo schema

$$A \xrightarrow{k_1} 2X \qquad X \xrightarrow{k_2} B \tag{4.23}$$

Chiaramente il primo passo equivale a creare un ammasso di particelle della specie X nel sistema. In altre parole, la cinetica chimica non-lineare dà luogo a un effetto che ricorda le interazioni attrattive. All'equilibrio, la proprietà del bilancio dettagliato assicura che ogni reazione, come $A \rightarrow 2X$, sarà in media tanto frequente quanto la sua inversa, $2X \rightarrow A$. Il sopra descritto raggruppamento verrà dunque ostacolato dal fatto che $2X \rightarrow A$ impoverisce il sistema delle coppie di particelle, giocando cosí un ruolo analogo alle interazioni repulsive. In questo modo verrà ripristinato il familiare comportamento poissoniano. Analogamente, se tutti i passi della reazione fossero lineari, le particelle verrebbero create o distrutte in posti casuali nello spazio, e ne risulterebbe ancora il prevalere di un comportamento poissoniano. La presenza sia della *cinetica non-lineare* che dei *vincoli di non-equilibrio* disturberà questo bilancio, e ne risulterà il mantenimento di una deviazione sistematica dalla legge di Poisson. Come si possono mettere in relazione queste deviazioni all'esistenza di correlazioni spaziali? Nel passo $A \rightarrow 2X$ le due particelle X sono ovviamente correlate, poiché sono state create insieme. Grazie alla diffusione, questa correlazione sarà estesa allo spazio finché almeno una delle particelle diventa inattiva attraverso il secondo passo dello schema, l'equazione 4.23. La lunghezza di correlazione risultante è dunque data dalla stima:

Tasso di inattivazione di X attraverso il secondo passo dell'equazione 4.23
$\simeq$ Tasso di diffusione su una distanza uguale alla lunghezza della correlazione

cioè

$$k_2 x \sim \frac{D}{l_{corr}^2} x$$

ovvero

$$l_{corr} = \sqrt{\frac{D}{k_2}} \tag{4.24}$$

dove D è il coefficiente di diffusione di Fick della specie X.

Si noti che se indichiamo con $F(X)$ la legge di produzione corrispondente allo schema 4.23, $F(X) = k_1 A - k_2 X$, k_2 è semplicemente uguale a

$|F'(X_s)|$ che, secondo il capitolo III, § 6, e l'*Appendice* A, dà il tasso di decadimento, γ, della perturbazione attorno allo stato stazionario $X_S = k_1 A/k_2$. Questa conclusione è in effetti del tutto generale: in ogni sistema soggetto a un processo di trasporto la cui velocità è data dal coefficiente D, e a un processo «di tipo chimico» il cui modo piú lento è caratterizzato dal tasso di decadimento γ, la lunghezza di correlazione è data da

$$l_{corr} \simeq \sqrt{\frac{D}{\gamma}} \tag{4.25}$$

Come abbiamo ripetutamente visto nel capitolo III, nelle vicinanze di un punto di biforcazione, γ tende a zero e ne risulta che la lunghezza di correlazione diverge. Questo indica a sua volta che la coerenza tende a includere l'intero volume del sistema, che diventa cosí in grado di subire collettivamente una transizione verso un nuovo stato. Si noti tuttavia che a causa del «rallentamento critico» dell'evoluzione vicino al punto di biforcazione, diventa impossibile discutere in termini di modi individuali e dei loro tassi di decadimento. È necessaria una nuova teoria piú completa che tenga conto dell'accoppiamento non-lineare dei modi. È stato elaborato un tale approccio, usando sia l'intera equazione principale sia il suo limite continuo. Le principali conclusioni qualitative sono che vicino al punto di biforcazione la legge della divergenza della lunghezza di correlazione è modificata rispetto all'equazione 4.25 a causa del modo di accoppiamento, e che la transizione può persino venire completamente soppressa.

Per comprendere quest'ultimo punto estremamente sorprendente, è necessario introdurre il concetto della *dimensionalità critica*. Ricordiamo che stiamo trattando con sistemi in cui un processo locale (come per esempio, una reazione chimica autocatalitica) coesiste con un processo di trasporto che accoppia fra loro i vicini elementi di volume (cfr. figura 4.3). A causa delle fluttuazioni, i processi di «tipo chimico» tenderanno a produrre deviazioni dalla omogeneità spaziale. Nel caso estremo del non accoppiamento fra elementi di volume, ciò dovrebbe portare a una sovrapposizione casuale di stati localizzati: anche se entro ogni volume i parametri sono vicini ai loro valori di biforcazione, il sistema non «vedrà» la biforcazione perché essa sarà sopraffatta dal rumore. D'altra parte, nel caso opposto in cui le celle spaziali sono accoppiate ma il processo di tipo chimico è assente, l'omogeneità spaziale sarà assicurata, ma l'ordine sarà nuovamente assente, poiché non ci sarà sempli-

cemente nessun «messaggio» che verrà ritrasmesso e mantenuto. Concludiamo dunque che sono necessari entrambi i processi e debbono essere di importanza paragonabile o, in termini piú concreti, debbono avere paragonabili tempi caratteristici. Ora il tempo caratteristico per appianare una disomogeneità attraverso il trasporto sopra una lunghezza L della cella spaziale è semplicemente dato da

$$\tau_D \simeq (D/L^2)^{-1}$$

o, usando la relazione $L^d = \Delta V$ dove d è il numero delle dimensioni dello spazio,

$$\frac{1}{\tau_D} \simeq D\, \Delta V^{-2/d} \tag{4.26}$$

e la coerenza spaziale viene stabilita sopra un raggio macroscopico, ΔV sarà grande, e quindi la diffusione sarà lenta. Inoltre, questo rallentamento sarà tanto piú importante quanto piú piccola è la dimensione dello spazio. D'altra parte, come abbiamo ricordato precedentemente, vicino a un punto di biforcazione il processo di tipo chimico sentirà un rallentamento critico, da cui risulterà che anche il suo tempo caratteristico τ_{car} diventerà grande. Tuttavia, in contrasto con τ_D il rallentamento sarà indipendente dalla dimensione, a causa del carattere locale del processo. Infatti, per una data cella ΔV, esso è dato dall'inverso del tasso di decadimento γ, del modo piú lento (cfr. equazione 4.23) e dipenderà dai parametri e dal grado delle non-linearità dominanti. Richiedendo che sia τ_D che τ_{car} siano dello stesso ordine, si arriva quindi a una relazione che lega il valore della dimensione alle proprietà intrinseche del sistema: questa è precisamente la dimensione critica, d_c. Per $d \gtrsim d_c$, la diffusione sarà efficace, e ne risulta che verrà mantenuta una correlazione spaziale a lungo raggio. Ma se d è molto inferiore a d_c, la diffusione non sarà in grado di correlare sufficientemente bene le varie celle spaziali e la biforcazione verrà «evitata».

Lo specifico valore di d_c dipende piuttosto debolmente dalla struttura particolareggiata del sistema, poiché ciò che sembra importare sono alcuni aspetti del tutto generali come il grado della non-linearità dominante. Si possono cosí definire «classi di universalità» costituite da sistemi completamente diversi, che tuttavia mostrano un comportamento qualitativo simile. Per esempio, sia nel modello autocatalitico dell'equazione 2.19 che nel brussellatore (equazione 3.18), la dimensione cri-

tica per alcune biforcazioni risulta essere $d_c = 4$. Si noti tuttavia che lontano dalla biforcazione la dimensione cessa di giocare un ruolo importante. Per esempio, la stima dell'equazione 4.25 per la lunghezza di correlazione rimane valida per qualunque valore di d.

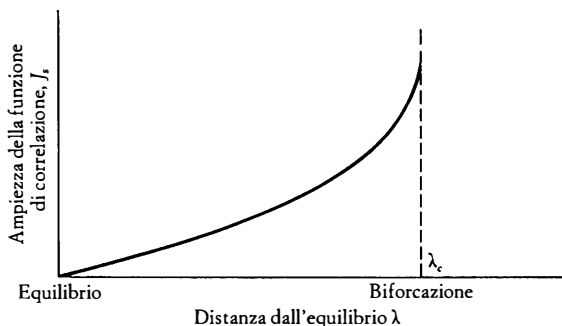
Per andare oltre questi risultati qualitativi dobbiamo costruire la forma esplicita delle soluzioni dell'equazione principale (equazione 4.12). Non discutiamo qui gli aspetti tecnici di questo problema. Schematizziamo alcuni punti importanti:

1. Le correlazioni di raggio macroscopico nascono improvvisamente non appena il sistema non è strettamente all'equilibrio, e possono essere mantenute indefinitamente. Nei sistemi in cui un modo di tipo chimico viene accoppiato a un modo di tipo trasporto, il raggio è intrinseco ed è essenzialmente dato dall'equazione 4.25. La forma precisa della correlazione dipende sia dalla geometria che dai parametri intrinseci. In un sistema tridimensionale di ampia estensione spaziale, troviamo dunque una funzione di correlazione tra le locazioni $\boldsymbol{r}$ e $\boldsymbol{r}'$ (strettamente legate alla covarianza, cfr. equazione 4.2) della forma

$$(4.27) \quad g(\boldsymbol{r}-\boldsymbol{r}') = \frac{J_s}{2\pi D} \frac{1}{|\boldsymbol{r}-\boldsymbol{r}'|} \exp\left[-\left(\frac{|F'(X_s)|}{D}\right)^{1/2} |\boldsymbol{r}-\boldsymbol{r}'|\right]$$

la cui ampiezza J_s è proporzionale alla forza del vincolo di non-

Figura 4.6

Non appena un sistema non è in equilibrio emergono le correlazioni spaziali a raggio macroscopico, la cui ampiezza è una funzione crescente dei vincoli di non-equilibrio.

equilibrio. In un certo senso, dunque, l'ampiezza J_s si può considerare come un «parametro d'ordine» che caratterizza la «transizione» a uno stato non correlato della materia (equilibrio) a uno stato correlato (non-equilibrio). D'altra parte nei sistemi in cui non c'è accoppiamento fra modi diversi, la lunghezza di correlazione può essere estrinseca. Per esempio, in un conduttore termico a riposo, soggetto a una differenza di temperatura, si trova che la lunghezza di correlazione è data essenzialmente dalla grandezza del sistema. L'ampiezza rimane tuttavia collegata alla forza del vincolo di non-equilibrio (in questo caso il flusso di energia attraverso il sistema) e gioca nuovamente il ruolo di «parametro d'ordine». La situazione è qualitativamente descritta nella figura 4.6.

2. Quando ci si avvicina a un punto di biforcazione la lunghezza di correlazione diverge, come detto piú sopra, ammesso che la dimensione non sia troppo bassa. Le correlazioni si estendono attraverso il sistema, anche se la loro lunghezza prima della biforcazione era intrinsecamente determinata. Questo dà luogo a un cambiamento qualitativo delle proprietà della sottostante distribuzione di probabilità, che è in qualche modo simile al cambiamento fra la figura 4.1*a* e 4.1*b*, la differenza è che ora si tratta con distribuzioni multivariate che dipendono dai valori delle variabili stocastiche X a ogni punto dello spazio $\mathbf{r}$. Piú in particolare, in un sistema infinitamente esteso come il modello autocatalitico dell'equazione 2.15 soggetto a una biforcazione a stati stazionari multipli senza rottura della simmetria spaziale, risulta che la distribuzione stazionaria di probabilità vicino alla singolarità cuspidale (cfr. figura 3.10*a*) ha la forma:

$$P_S(\{x_r\}) = \exp\left\{\frac{2\,\Delta V}{Q_S}\sum_r\left[\frac{(\lambda-\lambda_c)}{2}(x_r-x_S)^2-\frac{1}{4}(x_r-x_s)^4 + \frac{\mathfrak{D}}{8d}\sum_l(x_{r+l}-x_r)^2\right]\right\} \tag{4.28}$$

Qui $\lambda-\lambda_c$ è la distanza dalla biforcazione, è il tasso del passaggio fra vicini elementi di volume ΔV collegati con il coefficiente di diffusione di Fick D, Q_S una certa combinazione positiva dei parametri del sistema, x_r il valore locale della variabile di stato X_r

ridotta dell'elemento di volume ΔV ($x_r = X_r/\Delta V$). L'indice l segnala i vicini delle celle di r ai quali è limitato lo scambio di X attraverso i processi di trasporto presenti nel sistema.

Il carattere non gaussiano di questa distribuzione riflette la coerenza associata alla biforcazione: il sistema non può piú essere ripartito in tanti sottosistemi debolmente accoppiati. Risultati analoghi si ottengono per biforcazioni che portano a cicli limite e a strutture spaziali associate alla rottura di simmetria.

Una interessante manifestazione del venir meno del teorema del limite centrale è la dipendenza della varianza dalle fluttuazioni del numero di particelle in un volume finito di grandezza V. La valutazione di questa quantità dall'equazione 4.28 nel regime di stabilità marginale, a cui avviene la transizione verso un nuovo stato, porta a:

$$\langle \delta X^2 \rangle_V \sim V^{3/2}$$

Ciò è da contrapporre alla dipendenza di $\langle \delta X^2 \rangle_V \sim V$ che caratterizza il regime gaussiano prevalente al di sotto della biforcazione.

Si deve ricordare che l'esponenziale nell'equazione 4.28, noto come il potenziale di Landau-Ginzburg, è molto familiare dalla teoria delle transizioni di fase all'equilibrio. Usando questa analogia si possono calcolare molte proprietà rilevanti del sistema, come il raggio di correlazione, la varianza delle fluttuazioni, ecc., usando la tecnica del *gruppo di rinormalizzazione* sviluppato da Kenneth Wilson.

In breve arriviamo a una semplice, affascinante descrizione di come può nascere l'ordine in un sistema. In termini in qualche modo antropomorfici, l'ordine sembra essere un compromesso fra due antagonisti: il processo non-lineare di tipo chimico, che attraverso le fluttuazioni manda continuamente ma incoerentemente «segnali innovativi» al sistema; e il processo tipo trasporto che li cattura, li ritrasmette e li stabilisce. Disturbare il complicato bilancio fra questi due «attori» in competizione significa portare a tali cambiamenti qualitativi come a uno stato irregolare in cui ogni elemento del sistema agisce per conto suo, o, al contrario, uno stato «omeostatico» come un fossile in cui le fluttuazioni vengono schiacciate e viene imposta una piena uniformità.

La complessità sembra dunque essere limitata da entrambi i lati da due diversi tipi di stati di disordine. È difficile evitare l'impressione che questa conclusione, che segue da leggi del tutto generali che gover-

nano la dinamica dei sistemi fisico-chimici, fornisca chiavi preziose per dare dei modelli a sistemi che sono al di là del dominio della fisica e della chimica a livello di laboratorio. Il capitolo VI sarà dedicato al problema del «trasferimento di conoscenza».

5. *Il comportamento dipendente dal tempo delle fluttuazioni: la cinetica e l'ordine di tempo dell'autoorganizzazione.*

Alla fine del § 1 abbiamo identificato le due principali ragioni che motivano lo studio delle fluttuazioni nei sistemi dinamici non-lineari. La prima ragione, ovvero il meccanismo per il quale nasce e viene mantenuta la coerenza, ci ha portato alle analisi svolte nel § 4. La seconda, ovvero la cinetica delle transizioni fra stati, costituisce il soggetto di questo paragrafo.

Il nostro punto di partenza è l'equazione 4.15, che scriviamo nella forma piú generale

$$\frac{d\langle x\rangle}{dt} = \text{(legge di produzione macroscopica)} + \text{(deviazione dalla poissoniana)}. \qquad (4.29)$$

Come visto precedentemente, quando si incontra il punto di biforcazione la funzione di correlazione, e quindi la deviazione dalla poissoniana, subiscono un mutamento radicale a causa della divergenza della lunghezza caratteristica. A questo punto, dunque, si deve prevedere che le soluzioni dipendenti dal tempo dell'equazione principale 4.12 descrivano la crescita del modo critico sia nelle vicinanze che sopra l'instabilità. Con l'avanzare di questa crescita, la deviazione della varianza dal valore della poissoniana diventa sempre piú spiccato. Cosí, il secondo contributo al membro a destra dell'equazione 4.29 può non essere trascurabile in alcuni elementi di volume ΔV di piccola grandezza e, dopo un sufficiente (e generalmente molto lungo) lasso di tempo, esso può trasportare e guidare la media a un nuovo regime macroscopico.

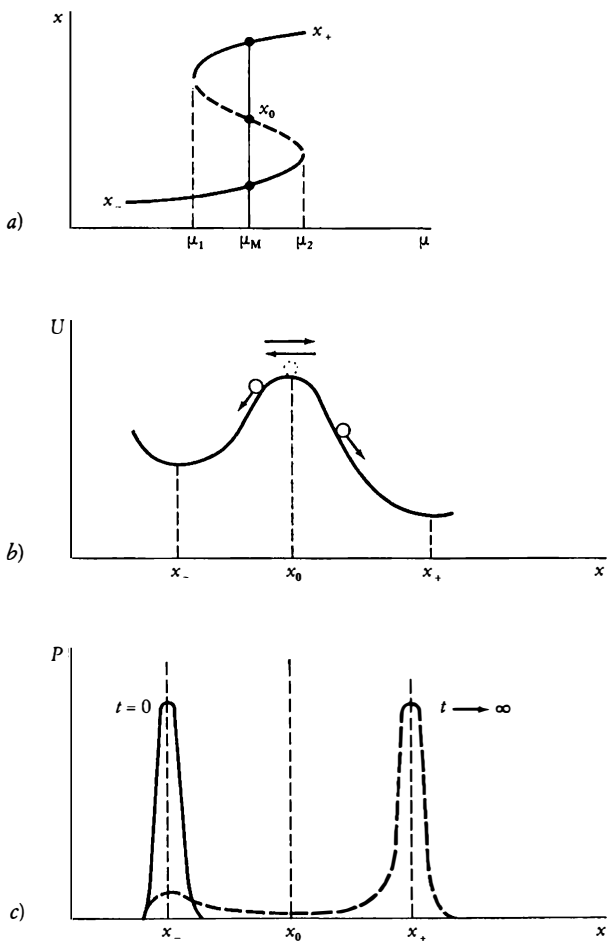
La figura 4.7 illustra questa *prospettiva teorica della nucleazione* delle transizioni di non-equilibrio, originariamente dovuta agli autori di questo volume. Il sistema studiato è il modello autocatalitico dell'equazione 2.15. Inizialmente si ha una probabilità di distribuzione con un picco attorno a uno dei rami (deterministicamente) stabili che si trovano dopo la biforcazione. La figura 4.7 fornisce tre aspetti alternativi di questa situazione.

Figura 4.7

Transizioni indotte dalla fluttuazione tra simultanei stati stazionari stabili.

a) Dipendenza delle soluzioni delle equazioni deterministiche dal parametro μ. Si è assunto che per $\mu = \mu_M$ lo stato x_- è meno stabile dello stato x_+.

b) Il potenziale cinetico U (cfr. equazione 4.30) come una funzione della variabile di stato x, per $\mu = \mu_M$. Le fluttuazioni casuali attorno ad ogni attrattore stabile x_+, x_- sono smorzate, a meno che esse non portino il sistema nell'altro bacino di attrazione attraversando la barriera di potenziale dello stato instabile $x = x_0$. La profondità piú grande di U in $x = x_+$ riflette l'ipotesi che x_+ è piú stabile di x_-.

c) Una distribuzione di probabilità inizialmente piccata in $x = x_-$ tenderà alla fine ad una forma bimodale la cui gobba in x_+ è molto piú larga che la gobba in x_-. La scala temporale di questa evoluzione è data dalla equazione 4.31.

Nella figura 4.7*a*, che va confrontata con la figura 3.10*b*, vi è il grafico degli stati stazionari multipli in funzione del parametro caratteristico μ – una opportuna combinazione delle costanti di produzione e delle variabili chimiche di controllo A e B. I rami superiore e inferiore x_+ e x_- rispettivamente sono asintoticamente stabili, mentre il braccio mediano è instabile. Nella figura 4.7*b* si trova il grafico di una quantità utile, sulla quale il lettore è stato brevemente edotto nel capitolo II, § 6. La quantità considerata è detta *potenziale cinetico* ed è definita attraverso la relazione

$$\frac{dx}{dt} = \frac{\partial U(x)}{\partial x} \qquad (4.30)$$

Nei sistemi che coinvolgono una sola variabile, U esiste sempre ed è uguale a meno l'integrale della legge di produzione: $U = -\int dx\ F(x)$. Nei sistemi a molte variabili, d'altra parte, è in generale impossibile garantire l'esistenza di un potenziale cinetico. In ogni caso, ogni qualvolta si può costruire, il potenziale risulta essere un'utile quantità poiché i suoi massimi e minimi corrispondono, rispettivamente, alle soluzioni asintoticamente stabili e instabili delle equazioni fenomenologiche di evoluzione. Infine, le linee piene nella figura 4.7*c* rappresentano la distribuzione di probabilità iniziale centrata sullo stato x_-.

Descriviamo ora i risultati principali. Supponiamo che il sistema possa essere mantenuto in condizioni di uniformità spaziale entro un volume ΔV che contiene N particelle della specie chimica attiva responsabile della bistabilità. Dalla nostra precedente discussione ci aspettiamo che la dimensione caratteristica di ΔV sia collegata alla lunghezza di correlazione. Una descrizione semplificata di questa situazione può essere basata sulla soluzione dell'equazione principale di un sistema omogeneo di grandezza N, in cui si trascura la diffusione. Si trova allora una transizione verso il secondo ramo stabile x_+ della figura 4.7*a*, il cui ordine di tempo è determinato da N e dalla differenza dei valori del potenziale U fra lo stato iniziale x_- e lo stato x_0. La distribuzione finale di probabilità viene rappresentata da linee tratteggiate nella figura 4.7*c*, per il caso in cui il minimo di U allo stato iniziale è molto piú basso del minimo di x_+.

Il meccanismo responsabile per il fenomeno sopra descritto è molto simile alla teoria dell'energia di attivazione della cinetica chimica, a cui si è alluso brevemente alla fine del § 4 del capitolo I. A causa delle fluttuazioni le diverse parti del sistema subiscono spontanee deviazioni dallo stato iniziale piú probabile x_-. La maggioranza di queste fluttua-

zioni hanno intensità e raggio piccoli, ma se la grandezza del sottosistema considerato non è molto grande, avverranno con una certa probabilità fluttuazioni di raggio finito. Alcune di queste fluttuazioni porteranno il sistema oltre la «barriera di potenziale» costituita dallo stato instabile x_0. A questo punto il sistema salterà velocemente all'altro stato attrattore x_+, costituendo quindi un «nucleo» per l'instabilità. Poiché per ipotesi lo stato x_+ è piú stabile dello stato x_-, la probabilità di scappare da esso sarà piccola finché nuovi elementi del sistema subiranno transizioni analoghe e alla fine l'intero sistema diventa contaminato.

Calcoli dettagliati mostrano che il tipico ordine di tempo per una transizione da x_- a x_+ è:

$$\tau_{trans} \simeq \exp N\,\Delta U \qquad (4.31)$$

dove $\Delta U = U(x_0) - U(x_-)$ è la «barriera di potenziale» che deve essere superata dalle fluttuazioni. Se un sistema fisico-chimico di ampia estensione spaziale, per il quale $N \sim 10^{23}$ viene mantenuto artificialmente uniforme, per esempio da una miscelazione estremamente efficiente, τ_{trans} sarà estremamente grande, molto piú grande di ogni concepibile tempo geologico o cosmologico. Questo mostra dunque che in situazioni reali la transizione deve partire come un evento localizzato, che coinvolge un elemento di piccole dimensioni. Anche cosí, il tempo caratteristico τ_{trans} rimarrà considerevolmente piú grande dei tempi caratteristici di rilassamento mostrati dalle leggi deterministiche di evoluzione a meno che il sistema non sia molto vicino ai punti di inversione μ_1 e μ_2 (figura 4.7*a*) in cui il valore della barriera di potenziale diventa molto piccolo.

La necessità di una descrizione locale sopra sottolineata implica che si debba ricorrere a una teoria dettagliata basata su una equazione principale multivariata che include gli effetti della diffusione (cfr. figura 4.3). Sono disponibili pochi risultati analitici; le simulazioni numeriche stabiliscono, tuttavia, che l'ordine di tempo della transizione dipende in modo molto interessante dal valore del coefficiente di diffusione. Quando questo coefficiente è piccolo in confronto alla produzione chimica, si trova che le transizioni da x_- a x_+, che inducono la comparsa di piccoli «nuclei» dello stato x_- entro lo stato x_+, avvengono ad alta velocità. Successivamente questi nuclei crescono lentamente e trascinano, per diffusione, il resto del sistema verso un nuovo ramo di soluzione x_+. Quando, d'altra parte, il coefficiente di diffusione è grande, si trova che lo stato iniziale x_- viene mantenuto per una considerevole quantità

di tempo. Questo ricorda la *metastabilità*, una proprietà familiare dalle transizioni di fase all'equilibrio. Come mostrato sperimentalmente da Roux e altri sulla reazione cloro-iodio, si osserva un simile comportamento quando il sistema è soggetto a perturbazioni disomogenee *esterne* che nascono, per esempio, dalle deviazioni da una miscelazione completamente efficiente.

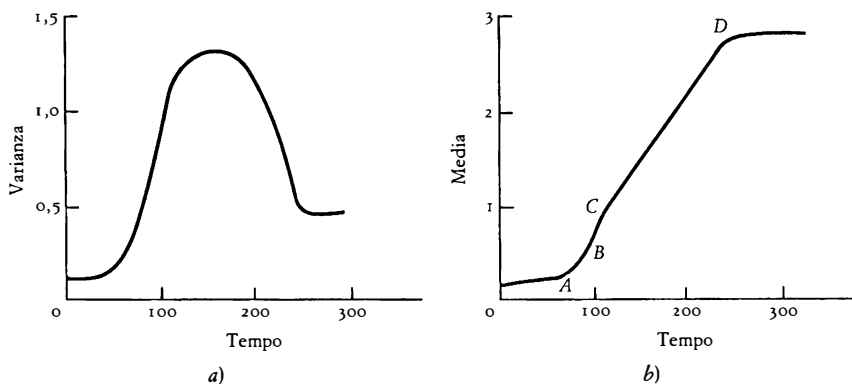
La figura 4.8 descrive la cinetica del valor medio (*b*) e della varianza delle fluttuazioni (*a*). Dopo un tempo di attesa relativamente lungo nasce il primo germe (punto A sulla curva della figura 4.8*b*) e un secondo segue al punto B. Al punto C i due germi raggiungono quasi simultaneamente i confini del sistema dopo di che si formano due fronti d'onda convergenti e alla fine scompaiono quando viene raggiunto lo stato D corrispondente al ramo x_+. Si noti l'apprezzabile aumento delle fluttuazioni attorno al punto di flesso nella curva del valor medio in funzione del tempo, come dimostrato dal brusco massimo della varianza.

In aggiunta ai fenomeni sopra descritti, ci si aspetta anche un ricco comportamento transiente nei sistemi in cui uno stadio di moto lento viene seguito da una evoluzione a scale di tempo molto piú veloci. Una possibile realizzazione di questo comportamento si ha quando il sistema viene preparato inizialmente a uno stato instabile o marginalmente stabile. Una seconda possibilità, alla quale ci limiteremo qui, nasce quando un *regime di induzione* iniziale caratterizzato da una velocità di cambia-

Figura 4.8

Transizione mediata da nucleazione fra stati stazionari multipli in un sistema modello.

a) Fluttuazione della varianza.

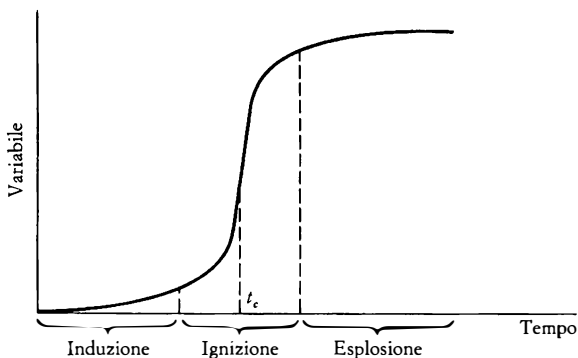
b) L'evoluzione temporale della media inizia a crescere fortemente non appena un germe di stato x_+, che appare inizialmente attraverso una fluttuazione, comincia a invadere una parte sostanziale del sistema, inizialmente nello stato x_-.

mento molto piccola della variabile pertinente, viene improvvisamente interrotto da un violento *comportamento esplosivo*, acceso per un qualche tempo caratteristico t_c. Ciò è illustrato nella figura 4.9. Alla fine il sistema si stabilizza a qualche stato finale, che per semplicità si assume essere l'unico stato stabile disponibile nel dominio di interesse dei valori dei parametri. Un esempio tipico in cui queste condizioni sono realizzate è la combustione. Ulteriori esempi comprendono reazioni chimiche autocatalitiche o fenomeni di commutazione nei laser.

Illustriamo ora gli aspetti principali di questo tipo di evoluzione sull'esempio della esplosione adiabatica, ovvero, della combustione in un recipiente chiuso nella quale il calore rilasciato dalle reazioni chimiche viene utilizzato per scaldare il miscuglio. Il caso piú semplice non banale consiste in un'unica reazione esotermica irreversibile

$$X \xrightarrow{k(T)} A \tag{4.32}$$

la cui velocità costante $k(T)$ dipende dalla temperatura T attraverso la legge di Arrhenius, $k(T) \simeq k_0 e^{-(E_0/k_B T)}$ (cfr. equazione 1.5). Le equazioni del bilancio della massa e della energia sono quindi date da

$$\begin{aligned} \frac{dx}{dt} &= -k(T)x \\ C_v \frac{dT}{dt} &= -r_v \frac{dx}{dt} = r_v k(T)x \end{aligned} \tag{4.33a}$$

Figura 4.9
Illustrazione di un comportamento esplosivo.

dove C_v, r_v sono rispettivamente il calore specifico e il calore di reazione a volume costante. Moltiplicando la prima equazione per *rv* e poi sommando, possiamo facilmente verificare che la seguente condizione di conservazione è soddisfatta per ogni valore del tempo:

$$C_v T_0 + r_v x_0 = C_v T + r_v x \equiv C_v T_{max} \tag{4.33b}$$

nella quale T_0, x_0 sono i valori iniziali di T e x; T_{max} la temperatura finale assunta quando la reazione viene completata. Grazie a questa legge di conservazione le equazioni 4.33*a* si possono chiudere. Per esempio, l'equazione del bilancio termico diviene semplicemente

$$\frac{dT}{dt} = (T_{max} - T) \cdot k(T) \tag{4.33c}$$

La figura 4.10 descrive la soluzione delle equazioni 4.33. Si vede che la velocità di reazione raggiunge bruscamente il suo valore massimo al tempo t_c, che verrà chiamato il «tempo di esplosione», in accordo con il quadro generale della figura 4.9.

Incorporiamo ora nella descrizione le fluttuazioni. Chiaramente stiamo considerando un processo di pura mortalità, poiché la concentrazione di X può solo decrescere. Possiamo pertanto scrivere un'equazione principale della forma dell'equazione 4.14, nella quale poniamo l'indice di natalità uguale a zero. Denotando con X il numero di particelle della specie reattiva, si ha

$$\frac{dP(X, t)}{dt} = \mu_{X+1} P(X + 1, t) - \mu_X P(X, t) \tag{4.34a}$$

L'indice di mortalità può essere facilmente calcolato grazie all'equazione di conservazione 4.33*b*:

$$\mu(X) = k(T)X = k_0 X \exp\left[-\frac{E_0}{k_B\left(T_{max} - \frac{r_v}{C_v}\frac{X}{N}\right)}\right] \tag{4.34b}$$

dove N è un parametro proporzionale alla grandezza del sistema.

La figura 4.11 raffigura gli stadi principali dell'evoluzione della distribuzione di probabilità prevista dalle equazioni 4.34. La condizione iniziale è scelta essere una distribuzione con un picco acuto centrato in uno stato che appartiene alla parte piatta (induzione) della figura 4.9,

per la quale la velocità di evoluzione è prevista essere molto piccola. Si osserva quindi un fenomeno di *bimodalità transiente* nel tempo: mentre uno dei massimi della distribuzione di probabilità rimane vicino allo stato iniziale, un secondo massimo si forma vicino allo stato di combu-

Figura 4.10

Soluzione delle equazioni 4.33. Curva *a*, soluzione per concentrazione, *x*; curva *b*, soluzione per temperatura, *T*; curva *c*, soluzione per velocità di reazione. Valori dei parametri: E_0/R = 10 000; T_{max} = 2000°; r_v/C_v = 1200.

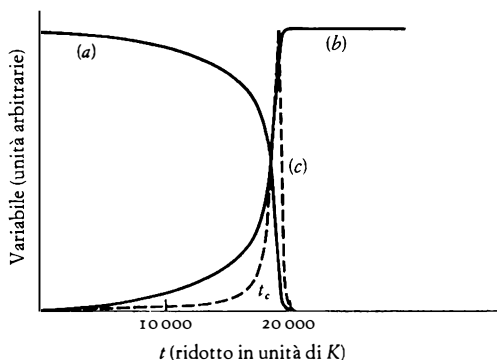
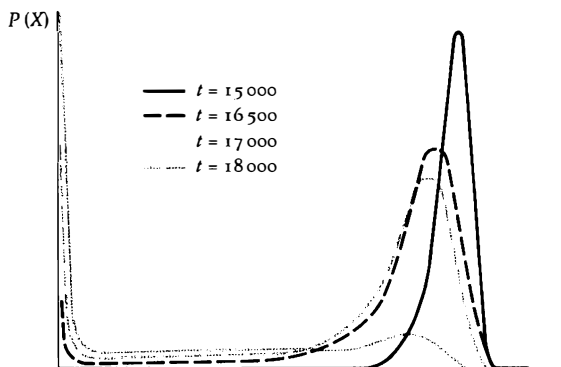



Figura 4.11

Bimodalità transiente che nasce in una esplosione adiabatica. Per t = 15 000 volte l'unità la probabilità è unimodale. Oltre t = 16 500 volte l'unità si sviluppa un secondo picco sempre piú pronunciato nella regione dei valori bassi del numero di particelle X. Nel frattempo l'ampiezza del picco iniziale, il cui moto segue la scala temporale lenta (con lenta induzione), diminuisce costantemente. Valori dei parametri come nella figura 4.10.

stione completa. Alla fine il primo picco scompare, e il sistema viene condotto al suo unico stato stabile. Abbiamo dunque un «comportamento di biforcazione nel tempo» durante l'evoluzione transiente del sistema.

Abbozziamo una spiegazione qualitativa di questo fenomeno inaspettato. Ricordiamo che stiamo trattando un processo che coinvolge due scale temporali largamente diverse, e che il nostro sistema viene inizialmente preparato in uno stato nel quale l'indice deterministico è molto basso. Il massimo della distribuzione di probabilità sottostante, il cui moto segue all'incirca quello deterministico, si muoverà quindi molto lentamente verso la regione di valori piú bassi di X. Nel frattempo, a causa delle fluttuazioni, la probabilità svilupperà una larghezza proporzionale alla lunghezza del periodo di induzione e inversamente proporzionale alla radice quadrata di N (per il teorema del limite centrale, cfr. § 1). Se la lunghezza del periodo di induzione è grande, l'effetto della grandezza viene ostacolato e la larghezza diventa apprezzabile. Come risultato, una parte sostanziale della massa di probabilità raggiungerà il punto di ignizione molto prima del massimo. A questo punto sarà rapidamente trascinato dal moto veloce verso la regione di bassi valori di X. Questa perdita di probabilità andrà avanti continuamente, ma poiché il sistema non può raggiungere valori negativi di X, nascerà un «ingorgo stradale» da cui risulterà un nuovo picco di probabilità nella regione delle X piccole. Alla fine il picco primario, adesso considerevolmente diminuito, raggiungerà il punto di ignizione e ciò marcherà la fine della bimodalità transiente.

Ovviamente, durante la vita della bimodalità transiente, il sistema mostrerà un comportamento marcatamente casuale: vi saranno apprezzabili deviazioni fra la media e i valori piú probabili, e la varianza raggiungerà un livello macroscopico. Ciò si mostrerà sotto forma di fluttuazioni del tempo di ignizione stesso, che diventerà quindi un evento casuale.

Oltre alle sue specifiche applicazioni a problemi correlati alla combustione, il fenomeno delle «biforcazioni nel tempo» fornisce un nuovo modello di differenziazione e di evoluzione nei sistemi al non-equilibrio, avendo origine puramente interna. In effetti, il sistema non necessita, per questo scopo, di essere continuamente turbato dal mondo esterno come, ad esempio, nella visione darwiniana dell'evoluzione attraverso le mutazioni e la selezione naturale. Piuttosto, le deviazioni della dinamica dall'equilibrio che si creano temporaneamente, sono sufficienti per indurre, durante un certo intervallo, una differenziazione

interna: durante il lento periodo dell'evoluzione, le fluttuazioni dotano il sistema di una varietà di stati con una certa probabilità, ma lo stato iniziale rimane dominante; l'apparire di uno stadio rapido trascina quindi alcuni di questi stati con una tale velocità che essi perdono traccia dell'«antenato» comune, producendo un proprio «clone» nuovo. Concetti analoghi sono utili anche in vari problemi collegati alla dinamica delle popolazioni e all'epidemiologia o, per tornare ai sistemi fisico-chimici in problemi che coinvolgono raggruppamenti, collassi e la proliferazione dei difetti nel processo della disorganizzazione o della frattura di un materiale.

6. *Sensibilità e selezione.*

L'analisi dei paragrafi precedenti ci fornisce alcuni strumenti utili per risolvere il problema della selezione fra stati, ripetutamente sollevato nel capitolo III. Supponiamo che un sistema possegga un numero di soluzioni simultaneamente stabili. Incorporando l'effetto delle fluttuazioni per mezzo dell'equazione principale, si possono calcolare le altezze dei picchi delle distribuzioni di probabilità, come anche il tempo mediamente necessario per lasciare la regione di attrazione di ogni stato stabile (cfr. equazione 4.31). Lo stato dominante – lo stato che verrà selezionato – sarà dunque lo stato per il quale il tempo di permanenza è maggiore, e attorno al quale vi è il piú alto picco di probabilità. Per quanto soddisfacente, questa discussione lascia tuttavia non risolto un problema chiave, ovvero, cosa succede quando i picchi di probabilità e i tempi di transizione di almeno due stati sono uguali.

Questa possibilità non deve essere accantonata come fosse un caso patologico. Alcuni dei fenomeni di transizione piú sorprendenti considerati fino ad ora in questa monografia corrispondono alla rottura delle simmetrie di vario tipo e dànno inevitabilmente origine a multipletti di stati che sono indistinguibili per quanto riguarda la stabilità e le fluttuazioni. Un esempio è la transizione di rottura della simmetria spaziale a livello macroscopico, discussa nel capitolo III, § 11. Un secondo esempio è la rottura della simmetria chirale a livello molecolare, della quale sono stati citati alcuni aspetti nel capitolo III, § 13. In ogni caso è necessario un nuovo meccanismo di selezione. In questo paragrafo descriviamo una tipica possibilità, ovvero, la selezione attraverso le interazioni di rottura di simmetria con i campi esterni.

La base di questo meccanismo è stata esposta nel capitolo III, § 4,

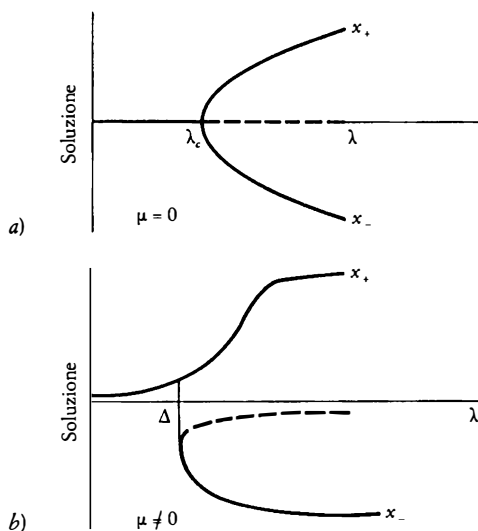
che abbiamo dedicato alla dipendenza dal parametro della biforcazione e al concetto collegato di stabilità strutturale. Come abbiamo rilevato in rapporto all'equazione 3.14 una biforcazione simmetrica che dà origine a due stati indistinguibili (cfr. figura 3.9*a* ed equazione 3.14 con $\mu = 0$) subisce un mutamento qualitativo allorché un parametro aggiuntivo cambia adeguatamente (cfr. figura 3.10*c* ed equazione 3.14 con $\mu = 0$). La situazione è riassunta nella figura 4.12. L'equazione corrispondente per la variabile pertinente, x, è

$$\frac{dx}{dt} = -x^3 + \lambda x + \mu \tag{4.35a}$$

e la sua soluzione per lo stato stabile è data da

$$-x_s^3 + \lambda x_s + \mu = 0 \tag{4.35b}$$

dove il parametro μ gioca il ruolo dell'asimmetria esterna. Anche se la

Figura 4.12
Selezione attraverso la perturbazione di una biforcazione per mezzo di un campo esterno μ che rompe la simmetria. La minima separazione tra i rami perturbati Δ, deve essere piú alta del livello di rumore prodotto dalla fluttuazione.

sua grandezza è piccola i suoi effetti possono essere amplificati considerevolmente data la natura non-lineare del sistema. Il ramo stabile x_-, che per $\mu = 0$ era indistinguibile da x_+ per quanto concerne la stabilità, esiste ancora, ma si separa dal ramo x_+ quando $\mu \neq 0$. Conseguentemente, al crescere del parametro λ dai valori a sinistra del punto limite oltre il quale x_- esiste, il sistema tenderà a stare sul ramo superiore della figura 4.12*b*, a meno che non vi sia una perturbazione di grandezza sufficiente a spingerlo nel ramo inferiore. Ora, poiché in ogni sistema si determinano fluttuazioni termiche, ciò può accadere se il Δ di separazione minimo della figura 4.12*b* è piccolo. Inversamente, se la simmetria esterna μ diviene un buon selettore, essa dovrebbe creare un Δ di separazione sufficientemente grande da superare le fluttuazioni termiche. Prima di discutere ulteriormente questo punto, tuttavia, vorremmo fare qualche osservazione sulla possibile origine del parametro della rottura di simmetria, μ.

Come rimarcato nel capitolo III, § 6, le equazioni come la 4.35 non sono limitate a particolari modelli, ma costituiscono piuttosto la tipica, *forma normale* che descrive intere classi di sistemi dinamici che funzionano vicino al punto di biforcazione. La questione che deve essere posta in connessione al parametro μ è, dunque, di identificare alcuni significativi campi esterni che dànno rottura di simmetria, i quali, se accoppiati a sistemi chimici, idrodinamici, ecc., dànno origine alla forma normale descritta dall'equazione 4.35. Menzioniamo qui solo due casi sulla base della loro universalità: il campo gravitazionale, che influisce su ogni oggetto dotato di massa sempre nella stessa direzione, e la corrente delle interazioni deboli-neutre che modifica leggermente le costanti chimiche di velocità di due reattivi stereoisomeri, sempre nella stessa direzione. In entrambi i casi si dimostra che la dinamica si può ridurre a una forma analoga all'equazione 4.35*a*, ammesso che in assenza della forzatura il sistema funzioni in vicinanza di un punto di biforcazione del tipo mostrato in figura 4.12*a*.

Veniamo ora a una stima quantitativa della sensibilità al piccolo parametro di rottura di simmetria μ. Dall'equazione 4.35*b* si vede facilmente che per $\mu \neq 0$ il punto imperturbato di biforcazione $\lambda_c = 0$ appartiene alla regione di soluzione unica, $x_s = \mu^{1/3}$. D'altra parte, la grandezza Δ della separazione minima si può ottenere dalla condizione che a questo punto l'equazione 4.35*b* abbia tre radici reali, due delle quali coincidenti. Si può facilmente vedere che (cfr. equazione 3.15):

$$\lambda_\Delta = 3(\mu^2/4)^{1/3} \qquad \text{e} \qquad \Delta = \frac{3}{2}(4\mu)^{1/3} \tag{4.36}$$

La presenza del fattore $\mu^{1/3}$ in questo risultato è di cruciale importanza. Se il sistema funziona vicino all'equilibrio (con λ ben al di sotto dello zero), l'effetto di un campo esterno è lineare, e dunque crea una piccola deviazione dell'ordine di $x_s \sim \mu$. Ma quando il sistema è lontano dall'equilibrio e gli è permesso manifestare pienamente la sua non-linearità, la stessa interazione crea ora una deviazione molto piú importante, $x_s \sim \mu^{1/3}$, che porta a una enorme amplificazione del segnale. Per esempio, nella rottura della simmetria chirale dove x_s indica la differenza fra stereoisomeri destri e sinistri, μ è la differenza dell'energia di attivazione (ridotta dall'energia termica) che sorge dalle interazioni deboli. Stime attendibili dànno $\mu \sim 10^{-15}$. Dunque all'equilibrio la deviazione ($x_s \sim 10^{-15}$) non è percettibile a livello macroscopico, ma lontano dall'equilibrio ($x_s \sim 10^{-5}$) può cominciare a competere con gli effetti macroscopici.

Come precedentemente menzionato, la precisione della selezione non si può valutare a meno che Δ non sia confrontabile con le fluttuazioni interne di x nelle vicinanze del punto di biforcazione. Le proprietà di queste fluttuazioni sono state discusse nei §§ 1 e 4. Per i nostri scopi, sono necessari solo alcuni aspetti generali. Specificatamente, dobbiamo prima confrontare il valore λ_Δ al punto di separazione minima delle soluzioni con il suo valore critico (biforcazione) $\lambda = 0$. Se λ_Δ è grande in confronto a N^{-1}, dove N è la grandezza del sistema, le fluttuazioni attorno al ramo x_+ saranno essenzialmente gaussiane e la loro varianza sarà

$$\sigma^2_{Gauss} \sim N^{-1} \tag{4.37a}$$

Ma se λ_Δ è molto piccolo, allora si deve calcolare la varianza usando la funzione quartica mostrata nell'equazione 4.28, che predice un valore dell'ordine di

$$\sigma^2 \sim N^{-1/2} \tag{4.37b}$$

La precisione della selezione implica $\Delta \gg \sigma$, quindi dalle equazioni 4.36 e 4.37 arriviamo alla condizione

$$\frac{3}{2}(4\mu)^{1/3} \gg N^{-1/2} \tag{4.38}$$

poiché questa disuguaglianza garantisce automaticamente che $\lambda_\Delta > N^{-1}$. Nel problema della rottura della simmetria chirale il lato sinistrorso è

dell'ordine di 10^{-5}. Per stimare il lato destrorso prendiamo una grandezza dell'ordine della lunghezza di correlazione (equazione 4.25) lungo la quale il sistema rimane spazialmente coerente e consideriamo una soluzione diluita in cui gli stereoisomeri attivi sono dell'ordine della millimole. Se si può raggiungere una lunghezza di correlazione di circa $l_{corr} \sim 1$ cm vicino alla biforcazione, allora $N = 10^{18}$ o $N^{-1/2} = 10^{-9}$, ed è soddisfatta la disuguaglianza 4.38. Si noti che vicino all'equilibrio essa si invertirebbe, poiché il lato sinistrorso sarebbe dell'ordine di $\mu \sim 10^{-15}$.

Avendo visto che la selezione è possibile in principio, per definire una buona strategia di selezione si dànno le seguenti prescrizioni:

1. Per $t = 0$, il sistema inizia vicino a $\lambda = \lambda_c = 0$, il punto di biforcazione in assenza del campo ($\mu = 0$). Si prende la distribuzione simmetrica e centrata per $x = 0$. Come già sottolineato, in questo dominio vi è un solo stato stabile, $x_s = \mu^{1/3}$. Il picco di probabilità verrà dunque gradualmente spostato verso questo stato. L'ordine di tempo si può stimare risolvendo l'equazione deterministica e dipendente dal tempo 4.35*a* per $\lambda = \lambda_c = 0$. Si ottiene con una semplice integrazione:

$$\Delta t_{0 \to x} \simeq \frac{1}{3\mu^{2/3}} \left[\ln \frac{\mu^{1/3}}{|x - x_s|} + \text{termini di ordine uno} \right] \qquad (4.39)$$

   Per x non troppo vicino a x_s l'ordine di grandezza di Δt è essenzialmente determinato dal fattore $\mu^{-2/3}$. Prendendo ancora l'esempio $\mu \sim 10^{-15}$, arriviamo a $\Delta t \sim 10^{10}$ sec $\sim 10^3$ anni! Quindi l'ordine di tempo deterministico per la selezione è esageratamente lungo.

2. Per evitare la difficoltà del precedente lungo tempo di selezione, lasciamo evolvere il sistema solo finché il picco di probabilità è migrato verso un valore x_{max} di x attorno al quale il livello di rumore (misurato dalla varianza σ) è piú piccolo di x_{max} stesso. In vista della nostra stima precedente, per $\mu \sim 10^{-15}$ sarebbe sufficiente raggiungere un valore di x_{max} leggermente superiore a 10^{-9}. A questo punto il sistema si smorza verso un nuovo valore del parametro λ, molto piú grande di λ_Δ. Essendo ora lontano dalle condizioni critiche, il sistema evolverà in modo esponenzialmente veloce verso il ramo x_+, molto prima che le fluttuazioni lo portino a uno stato x_- sostanzialmente popolato. Ci si aspetta che la selezione sia quasi perfetta. Le simulazioni numeriche mostrano

che con questa procedura la scala del tempo si accorcia a pochi giorni, invece dei 10^3 anni previsti dalla equazione deterministica.

Ancora una volta proviamo una sensazione di sbalordimento di fronte alle stupefacenti potenzialità della materia quando sono presenti sia la dinamica non-lineare che i vincoli di non-equilibrio.

7. *Dinamica simbolica e informazione.*

Fino ad ora abbiamo descritto la dinamica stocastica in termini di distribuzioni di probabilità e i loro vari momenti. Un approccio complementare e per molti scopi particolarmente illuminante, consiste nello studio delle realizzazioni individuali dei processi stocastici di interesse. In questo paragrafo riassumiamo alcuni elementi fondamentali di questo approccio, che saranno ulteriormente utilizzati nel § 8 e nel capitolo v.

Siano $\{Q_i\}$ ($i = 1, 2, \ldots$) gli stati accessibili di un sistema[1]. In accordo con le osservazioni fatte nel § 2, è chiaro che l'insieme $\{Q_i\}$ viene scelto in modo tale che il tempo di evoluzione definisce un processo di Markov. Per di piú, invece di considerare il tempo come un parametro continuo, assumiamo che i salti fra i vari stati avvengano agli istanti $t_m = m\,\Delta t$, dove m è un intero e Δt ha un valore fissato che dipende dal particolare processo in considerazione. Ciò non costituisce una assunzione molto restrittiva, poiché in natura le successive transizioni debbono in effetti essere distanziate da qualche intervallo di tempo. Naturalmente, Δt deve essere inferiore di molto alla scala di tempo caratteristica dell'evoluzione del sistema descritto dalle equazioni fenomenologiche del bilancio.

Nelle precedenti condizioni ogni possibile risultato di una serie di ρ passi consecutivi del processo si può scrivere come una sequenza

$$Q_{i_1}\, Q_{i_2} \ldots Q_{i_\rho} \tag{4.40}$$

L'insieme di tutte le sequenze della forma 4.40 definisce una *catena di Markov*. Guardare al sistema come a una catena di Markov equivale dunque a descrivere la sua evoluzione in termini delle «lettere» $\{i_l\}$ di

[1] Consideriamo dapprima un sistema con uno spazio degli stati discreto. Il caso delle variabili continue viene brevemente discusso alla fine del presente paragrafo.

un opportuno «alfabeto». A prima vista sembra che per la maggior parte dei sistemi di interesse il numero delle lettere nell'alfabeto debba essere infinito. Ma non è necessariamente cosí. Come esempio, consideriamo un sistema che ammette due stati stazionari simultaneamente stabili. A causa delle transizioni fra stati indotti dalle fluttuazioni (cfr. figura 4.7), il sistema trascorrerà una considerevole quantità di tempo vicino a questi stati, la cui durata sarà in media data dall'equazione 4.31, interrotta da una veloce transizione verso l'altro stato. Introduciamo una descrizione piú astratta in cui lo stato istantaneo è contraddistinto da «0» e «1», a seconda che il sistema sia nel dominio di attrazione del primo o del secondo stato stazionario stabile. Un osservatore che segue l'evoluzione notando la regione di attrazione in cui il sistema si trova a intervalli di tempo regolari vedrà dunque una sequenza del tipo

$$111001100010110\ldots \tag{4.41}$$

che fornisce una descrizione in termini di un alfabeto che coinvolge solo due lettere, 0 e 1. In principio, sequenze del tipo 4.41 non definiscono una catena di Markov. Tuttavia, sotto certe condizioni si può mostrare che la partizione dello spazio degli stati iniziali (che coinvolge, tipicamente, un'infinità numerabile o anche un continuum di stati) nei due stati di cui sopra, conserva il carattere di Markov. Si parla allora di *partizione di Markov* (*markovian coarse graining*). La descrizione permessa da sequenze del tipo dell'equazione 4.41 si chiama *dinamica simbolica*: non si ha piú a che fare con la successione degli stati nel senso tradizionale del termine ma, piuttosto, con una successione di simboli – le lettere di un alfabeto di lunghezza finita.

Proviamo a caratterizzare l'abbondanza di diverse sequenze del tipo dell'equazione 4.40 o 4.41. Dalla definizione di processo di Markov la probabilità dell'evento descritto dall'equazione 4.40 è

$$P(C) = P_{i_1} p_{i_1 i_2} \cdots p_{i_{p-1} i_p} \tag{4.42}$$

dove P_{i_1} rappresenta la distribuzione di probabilità iniziale dello stato Q_{i_1} e $p_{i_k i_{k+1}}$ la probabilità di transizione dallo stato Q_{i_k} allo stato $Q_{i_{k+1}}$ durante l'intervallo Δt. Si noti che, tipicamente, un dato stato verrà incontrato molte volte durante il processo.

Si può ora dimostrare il seguente teorema: sia di Markov il processo descritto dalla sequenza in considerazione e soddisfi alla legge dei gran-

di numeri (cfr. § 1). Il numero di tutte le sequenze di lunghezza ρ che coinvolgono N stati è chiaramente $N_\rho = N^\rho$. Disponiamo queste sequenze in ordine decrescente della probabilità $P(C)$ e selezioniamo, in questo ordine, quelle sequenze la cui somma delle probabilità eccede un prefissato numero positivo fra 0 e 1. Allora, nel limite di grandi ρ, il numero n_ρ delle sequenze cosí selezionate è indipendente dal numero prefissato e si può stimare come:

$$n_\rho \sim e^{\rho I} \qquad (4.43a)$$

dove

$$I = -\sum_{ij} P_i p_{ij} \ln p_{ij} \qquad (4.43b)$$

La quantità I è l'entropia della catena di Markov e viene generalmente vista come una misura della quantità di informazione ottenuta quando la catena di Markov si muove di un passo in avanti a partire dallo stato iniziale Q_i. La connessione fra questa quantità e la definizione di entropia usata nell'equazione 4.16 è chiara. È sufficiente considerare una probabilità iniziale «deterministica» ($P_i = 1$ se $i = i_1$, $P_i = 0$ altrimenti), nel qual caso $p_{i_1 j} = P_j$ rappresenta la probabilità di essere nello stato Q_j.

La grande importanza del precedente teorema si può comprendere dalle seguenti considerazioni. Mentre, come appena detto, il numero di tutte le sequenze del tipo (C) è $N_\rho = e^{\rho \ln N}$, il numero n_ρ delle sequenze selezionate è approssimativamente $e^{\rho I}$. Ora, possiamo richiamare (cfr. § 3) il fatto che l'entropia di un processo di Markov assume il suo massimo valore quando tutti gli stati sono ugualmente probabili, $P_i = 1/N$. Questo dà $I_{max} = \ln N$. Cosí abbiamo sempre $I < \ln N$ (tranne per il caso banale) e ne risulta, per ρ grandi,

$$e^{\rho I} \ll e^{\rho \ln N} \qquad (\rho \to \infty)$$

In altre parole, solo una piccola frazione di tutte le sequenze C possiede una somma di probabilità sufficientemente vicina all'unità. Quanto piccola sia questa frazione dipende dall'entropia della catena di Markov. Arriviamo cosí all'importante conclusione che una deviazione dall'equiprobabilità agisce come un selettore estremamente efficiente. Una tale deviazione si può realizzare se il sistema non è isolato ma viene invece messo in contatto con un ambiente esterno. Come abbiamo ripetutamente visto, tali condizioni possono portare a *stati di non-equilibrio*

che possono essere mantenuti indefinitamente grazie allo scambio di materia ed energia fra il sistema e il mondo esterno. Il non-equilibrio è dunque l'ambiente naturale in cui ha luogo la selezione.

Come esempio, consideriamo il processo di formazione di un biopolimero, per esempio una proteina lunga $\rho = 100$ amminoacidi. È ben noto che in natura esistono $N = 20$ tipi di amminoacidi. Se la biopolimerizzazione dovesse aver luogo in un sistema isolato in cui tutte le sequenze sono ugualmente probabili *a priori*, si avrebbe $N_\rho = e^{100 \ln 20} \sim e^{300}$ sequenze ugualmente probabili. Quindi, ogni particolare arrangiamento – necessario per adempiere, per esempio, a una desiderata funzione biologica – si verificherebbe con la straordinariamente bassa probabilità di e^{-300}! Se, d'altra parte, la biopolimerizzazione viene eseguita sotto condizioni di non-equilibrio corrispondenti a una entropia di, per esempio, $I = 0{,}1 I_{max}$, allora solo $n_\rho \sim e^{30}$ si possono realizzare sequenze di lunghezza 100 con apprezzabile probabilità. Questo numero è già molto piú vicino alla nostra scala. È cosí concepibile che l'evoluzione che agisce su una tale popolazione relativamente ristretta in un ambiente di non-equilibrio possa produrre, dato un certo tempo, entità dotate di speciali proprietà come l'autoriproduzione, trasduzione efficiente di energia, ecc.

Fino ad ora in questo paragrafo abbiamo assunto che i sistemi dinamici che dànno origine al processo stocastico in esame si possano definire in uno spazio discreto in cui l'evoluzione era un processo di Markov. Come abbiamo spiegato nel primo paragrafo di questo capitolo, queste condizioni possono essere soddisfatte dalle fluttuazioni interne generate spontaneamente dalla dinamica molecolare di tutti i sistemi fisicochimici macroscopici. Ma cosa si può dire della dinamica caotica, la seconda sorgente universale di stocasticità in natura? Per cominciare, lo spazio degli stati dei sistemi che dà origine a una tale dinamica è, tipicamente, continuo. Inoltre, non c'è nessuna ragione *a priori* per avere un processo di Markov, poiché, contrariamente alle fluttuazioni, le deviazioni dal moto medio non sono eventi né localizzati, né di piccola scala.

È notevole come in molti casi la dinamica caotica possa nonostante tutto essere sistemata nella forma della catena di Markov. La ragione è che lo spazio degli stati si può frequentemente ripartire in un numero finito di celle tali che, col passare del tempo, la dinamica induca transizioni fra queste celle che soddisfano la proprietà di Markov. Grazie alla esistenza di queste partizioni si può raggiungere un nuovo livello di astrazione in cui la *dinamica simbolica* che coinvolge la successione delle lettere di un alfabeto che segnala le varie celle diventa la descrizione na-

turale. Nel prossimo paragrafo discuteremo un particolare esempio di una tale descrizione simbolica, nel contesto del problema della selezione di strutture asimmetriche.

8. *Genesi di strutture asimmetriche e ricche di informazione.*

Abbiamo visto come il valore dell'entropia di una catena di Markov condizioni il numero delle sequenze che possono avvenire con una notevole probabilità, rivelando quindi delle informazioni. Inoltre, grazie all'uso delle dinamiche simboliche, è possibile considerare queste sequenze selezionate come un insieme di «messaggi» o «testi» contenenti informazione. Il problema che vogliamo sollevare in questo paragrafo è come questi messaggi possono essere generati spontaneamente da un sistema dinamico.

Quando abbiamo un pezzo di carta contenente un testo scritto in italiano, la nostra comprensione del messaggio dipende dal fatto che lo si legga da sinistra a destra, che si osservi la punteggiatura, le virgole e le regole sintattiche; un testo scritto in arabo avrà senso solo se è letto da destra a sinistra, e se si rispettano ancora le regole di sintassi che governano la lettura (in generale diverse da quelle italiane); analogamente, come osservato nel capitolo III, § 13, il codice genetico è caratterizzato da segnali di partenza che permettono di leggerlo in una direzione fissa dal punto di partenza. Un secondo aspetto, ugualmente importante, nella comprensione di un messaggio è la presenza di un particolare insieme di simboli che si dispiegano nel tempo, o, piú in generale, nella direzione della lettura. Questa sequenza, sebbene perfettamente definita e riproducibile una volta conosciuta, è fondamentalmente impredicibile, nel senso che in tutti i casi non banali le sue strutture globali non possono essere inferite dalla conoscenza di parte di essa, per quanto estesa. Sotto questo aspetto si può considerare come un processo stocastico, *ed è questa identificazione che ci consente di parlare di informazione.* Ad esempio ciò che noi percepiamo come informazione derivante dalla lettura degli *Elementi* di Euclide, o dei *Principia* di Newton sarebbe molto diversa se fossimo in grado di ricavare, attraverso un semplice algoritmo, la seconda metà di uno di questi trattati dalla lettura della prima! In breve l'informazione prevede due condizioni fondamentali:

1. Una netta rottura di simmetria nello spazio, a cui si deve la costante eliminazione di altri possibili esiti del processo di lettura

(questo ci riporta ai problemi sollevati nel capitolo III, § 13, e sopra, § 6, in particolare a come tali forme asimmetriche possano nascere e dominare la loro immagine speculare).

2. Un elemento di impredicibilità associato alla rivelazione di un oggetto o di un messaggio che il lettore non è in grado di inferire dall'inizio.

Abbozzeremo ora un meccanismo in cui entrambe queste condizioni sono automaticamente soddisfatte.

Consideriamo un sistema dissipativo dinamico, come un insieme di reazioni chimiche che diano origine alla sintesi di specie la cui concentrazione indichiamo con X_i. Supponiamo di registrare l'insieme di valori di alcune variabili rappresentative X_i ottenuti per tempi sempre piú lunghi. Possiamo considerare questa sequenza come una stringa modimensionale di cifre. Ora, a causa dell'irreversibilità inerente all'evoluzione di ogni sistema dissipativo, la sequenza «diretta» cosí generata sarà asimmetrica, nel senso che essa sarà diversa da una sequenza che potremmo leggere in direzione inversa: quest'ultima potrebbe di fatto essere generata soltanto da un sistema dinamico fisicamente inaccettabile risultante da un'inversione temporale iniziale. Vediamo quindi come l'irreversibilità può indurre un'asimmetria assoluta nello spazio astratto delle sequenze simboliche monodimensionali. Affronteremo ora il problema di come trasporre questa asimmetria di dinamica simbolica a un sistema monodimensionale incluso nello *spazio fisico*.

Proporremo quindi un algoritmo particolare grazie al quale raggiungere il nostro scopo. Assumiamo che il sistema chimico considerato sia in grado di produrre oscillazioni prolungate nel tempo, oscillazioni che non sono necessariamente periodiche. Assumiamo inoltre che quando X_i incrocia un certo livello L_i con una pendenza positiva (nei punti segnati • nella figura 4.13), si avvia un nuovo processo, il cui risultato è che la sostanza *i* precipita o diffonde al di fuori dello spazio della reazione, e viene successivamente raccolta su di un «nastro». La struttura spaziale che ne risulta porterà un calco delle varie soglie incontrate dalle diverse concentrazioni. Come si è notato sopra, questa sequenza sarà tipicamente asimmetrica a causa della rottura di simmetria temporale data dall'irreversibilità. Siamo dunque riusciti a disegnare una mappa dell'irreversibilità su di una struttura spaziale monodimensionale. Piú in generale si può mostrare che in un sistema di reazione-diffusione accoppiato a un vettore *polare* (come il campo gravitazionale o un campo elettrostatico), possono essere generate strutture spaziali che recano il

calco della sequenza di valori irreversibilmente toccati dalle variabili di concentrazione nel corso del tempo. Questa mappa spaziale dell'irreversibilità non può essere costruita in relazione a un vettore assiale, come un campo magnetico, poiché tale accoppiamento non può apparire nell'equazione di equilibrio di una variabile scalare come la concentrazione. Per questo motivo il meccanismo discusso qui non dà informazioni sulla selezione della chiralità, la seconda grande forma di asimmetria osservata in natura.

Classifichiamo ora le strutture che possono essere generate da diversi processi chimici. Supponiamo dapprima che $\{X_i\}$ tenda asintoticamente verso un ciclo limite. Sappiamo dal capitolo III che questo tipo di comportamento richiede almeno due variabili accoppiate, e che la differenza di fase tra due qualsiasi di esse sarà costante per ogni valore del tempo. Ne segue che i passaggi di livello al piú genereranno delle strutture definite dalla seguente dinamica simbolica, se due variabili arrivano a superare i livelli assegnati:

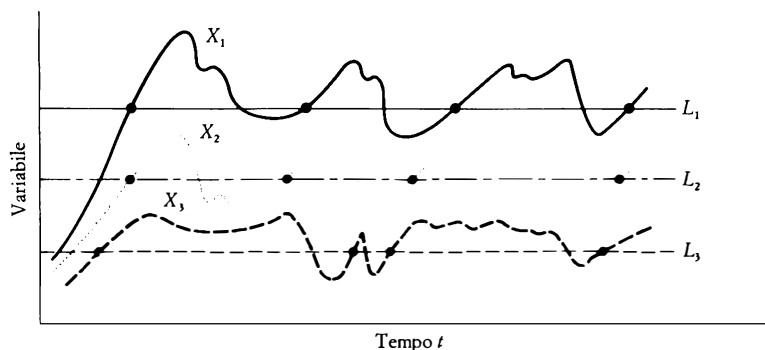
a) ...XYXYXY...

o

b) ...XYZXYZXYZ...

se tre variabili superano i livelli assegnati, e cosí via.

Figura 4.13

Generazione di stringhe di simboli da un sistema dinamico. Quando le variabili di stato X_1, X_2, ... attraversano i livelli L_1, L_2, ... con pendenza positiva avviene un processo di precipitazione, che permette il trasferimento di questa sequenza temporale lungo una coordinata spaziale.

La sequenza *a*) è reversibile, ovvero si legge nello stesso modo in entrambe le direzioni. Ma la sequenza *b*) descrive un oggetto assolutamente asimmetrico. Introducendo un simbolo abbreviato (l'analogo di un *codon*)

$$\alpha = \{XYZ\}$$

la sequenza *b*) diviene

c) $\ldots\alpha\alpha\alpha\ldots$

e perde in questa forma la sua asimmetria spaziale.

Sia che la struttura vada intesa in termini di *b*) o *c*), difficilmente si può parlare di informazione, poiché il sistema è completamente predicibile. Per esempio, applicando la definizione di informazione di Claude Shannon (equazione 4.43*b*), $I = -\sum_i P_i \ln P_i$, allo schema (*c*), troveremo che I è uguale a zero.

Supponiamo ora che il sistema continui a funzionare su un attrattore caotico. Come sottolineato nel precedente paragrafo, e come vedremo ancora nel capitolo v, sotto determinate condizioni l'evoluzione temporale di un sistema dinamico instabile può essere rappresentato da un processo stocastico per mezzo di una appropriata dinamica simbolica. In questo senso la trascrizione di una dinamica caotica secondo il meccanismo esposto sopra è potenzialmente in grado di dare origine a strutture che sono

1. asimmetriche, a causa dell'irreversibilità; e
2. ricche di informazione, poiché l'entropia di informazione teoretica di un processo stocastico è generalmente diversa da zero.

Illustriamo questa congettura con i risultati di simulazione numerica usando il modello di Rössler (equazione 3.30). Per i valori dei parametri $a = 0{,}38$, $b = 0{,}3$ e $c = 4{,}5$, questo modello dà origine a un attrattore caotico descritto nella figura 3.26. Un livello ragionevole di valori di soglia risulta essere $L_x = L_y = L_z = 3{,}0$. Usando le condizioni iniziali $X_0 = Y_0 = Z_0 = 1{,}0$, possiamo utilizzare il meccanismo del passaggio di livello descritto sopra, per generare una sequenza di simboli – tenendo conto che la stampa parte solo dopo che è trascorso un tempo sufficientemente lungo perché i transienti si estinguano – della forma:

ZYXZXYXZXYXZYXZXYXZYXZYXZXZYXZYXZXYXZYX...

Si può verificare che questa sequenza può essere riformulata introducendo gli «ipersimboli»

$$\alpha = ZYX, \qquad \beta = ZXYX, \qquad \gamma = ZX$$

Ne risulta dunque:

$$\alpha\beta\beta\alpha\beta\alpha\alpha\gamma\alpha\alpha\beta\alpha\ldots$$

Questa interessante proprietà va attribuita all'origine deterministica del meccanismo che genera le sequenze. Essa suggerisce l'esistenza di una forte correlazione nella successione dei simboli X, Y, Z, ovvero un processo di Markov di ordine elevato. Per verificare ciò valuteremo come segue le proprietà statistiche delle sequenze. Siano le equazioni 3.30 integrate fino a circa 40 000 unità temporali, generando cosí qualcosa come 10 000 simboli e 3000 ipersimboli. I risultati vengono registrati a partire da $t = 300$ perché i transienti si estinguano. Si contano quindi i numeri dei simboli singoli (X, Y, o Z), dei doppietti (XX, XY ecc.), tripletti, e cosí via. Le probabilità condizionali si deducono dunque attraverso relazioni come

$$P(B|A) = \frac{P(AB)}{P(A)}$$
$$P(C|AB) = \frac{P(ABC)}{P(AB)} \tag{4.44}$$

dove A, B, C possono essere uno qualsiasi di X, Y, e Z[2]. Il fatto che i doppietti come XX non occorrano mai mostra che la sequenza non è completamente casuale. Inoltre, l'assenza del tripletto YXY esclude la possibilità che la sequenza definisca un semplice processo di Markov, poiché in tal caso

$$P(\mathrm{YXY}) = P(\mathrm{Y})P(\mathrm{X}|\mathrm{Y})P(\mathrm{Y}|\mathrm{X})$$

che sono tutti diversi da zero.

Dobbiamo dunque guardare a un processo di Markov di ordine piú alto, nel quale la memoria di un evento si estende a molti passi indietro.

[2] Aumentando il tempo di integrazione possiamo constatare che la probabilità di trovare un determinato simbolo o successione di simboli tende a un limite definito. Sembra pertanto che abbiamo a che fare con un processo stocastico stazionario.

Si raggiunge un buon accordo con le frequenze dei simboli calcolate numericamente assumendo un processo di Markov del quinto ordine[3]:

$$\begin{aligned} P(ABCDEFG) &= P(ABCDEF)P(G|BCDEF) \\ &= P(ABCDE)P(F|ABCDE)P(G|BCDEF) \end{aligned}$$

dove $P(G|BCDEF)$ rappresenta la proprietà condizionale calcolata numericamente dell'ultimo simbolo, dato il quintupletto *BCDEF*.

Un esame dettagliato delle proprietà statistiche delle sequenze rivela alcune caratteristiche sorprendenti. Ad esempio, di tutte le possibili 3^7 sequenze di sette simboli che possono essere costruite a partire da X, Y, e Z, solo 21 sono quelle effettivamente realizzate dalla dinamica. Inoltre per circa la metà di esse la probabilità condizionale dell'ultimo simbolo, dati i cinque precedenti, risulta essere uguale a 1. Pertanto tutto accade come se il sistema fosse dotato di un insieme di regole «grammaticali», che vengono seguite automaticamente, come risultato della dinamica.

La procedura descritta può essere ripetuta cambiando il valore del parametro *a* nel modello (cfr. l'equazione 3.30) in 0,40. I risultati sono qualitativamente simili, compresa la possibilità di esprimere le sequenze in termini di ipersimboli, eccetto che per un mutamento nei valori numerici delle probabilità dei differenti multipletti. Analogamente un piccolo cambiamento nei valori di soglia lascia immutati i risultati principali. Un cambiamento di maggiori dimensioni nei valori di soglia ha tuttavia effetti drastici. Ad esempio, usando il tempo medio di X, Y e Z come soglia per queste tre quantità, troviamo sequenze che non possono essere espresse interamente in termini di ipersimboli. Inoltre un modello di catena di Markov del sesto ordine non è in grado di descrivere i dati adeguatamente.

Il contenuto di informazione della sequenza di X, Y e Z può essere calcolato da un'estensione della definizione di Shannon (cfr. equazione 4.43*b*) in cui lo stato *i* è ora espresso da un quintupletto di simboli. Si trova che esso è finito, ma molto inferiore al massimo, corrispondente all'equiprobabilità. Producendo queste sequenze come un naturale risultato della sua dinamica di non-equilibrio, a partire dalle numerose sequenze casuali che si possono prospettare, il sistema è cosí diventato

[3] Che la catena debba essere almeno del quarto ordine si può inferire dall'esistenza dell'ipersimbolo piú lungo, β, e dal fatto che *Z* è sempre preceduto da *X*.

un generatore di informazione. Si noti comunque, che ogni messaggio individuale rimane considerevolmente impredicibile, come testimoniato dal carattere stocastico della sequenza. D'altro canto il grado di impredicibilità è attenuato dall'esistenza di correlazioni che si estendono sui cinque vicini piú prossimi della catena. Prendendo a prestito un termine usato nella teoria della comunicazione, possiamo dire che la dinamica caotica ha reso possibile il *condensare* l'informazione contenuta potenzialmente nelle sequenze casuali riproducendole piú sinteticamente in termini di ipersimboli α, β e γ. Ritroviamo cosí, nel nostro semplice modello, uno degli aspetti essenziali dell'informazione rilevati all'inizio del paragrafo.

Riassumendo: all'inizio ci troviamo di fronte alla difficoltà che la stessa esistenza dell'informazione solleva due aspetti che sembrano essere in contraddizione fra loro. Da un lato l'informazione deve essere associata a una qualche forma di casualità (da qui i numerosi tentativi di collegarla con l'«ignoranza» e con l'«incompleta specificazione» di un sistema). Dall'altro, tuttavia, il numero astronomicamente grande di sequenze casuali di lunghezza $\rho \gg 1$ fa sí che sia estremamente improbabile selezionare, su base aprioristica, la particolare classe di sequenze che gioca probabilmente il ruolo principale in un fenomeno osservato.

Abbiamo suggerito un meccanismo che ci consente di realizzare un compromesso tra queste restrizioni: strutture informativamente significative possono essere generate da un sottostante meccanismo nonlineare, temporalmente irreversibile, e operante nella regione caotica lontana dall'equilibrio. In questo modo casualità e asimmetria – due requisiti necessari all'informazione – sono dall'inizio incorporati nella struttura risultante. In piú, poiché queste strutture sono il risultato di un meccanismo (che in definitiva ha a che fare con l'esistenza di molecole dotate di adeguate proprietà catalitiche), esse superano automaticamente la difficoltà della tremenda «improbabilità termodinamica» che caratterizza le sequenze casuali.

Uno degli aspetti piú significativi dei concetti discussi in questo e nel precedente paragrafo è l'emergere di un *nuovo livello di descrizione* rivelato dalla sottostante dinamica. Cosí come diviene naturale descrivere lo stato di un sistema non in equilibrio in termini delle correlazioni tra elementi macroscopicamente separati piuttosto che delle forze intermolecolari, e descrivere il fenomeno della biforcazione in termini del parametro d'ordine piuttosto che di una delle variabili di stato inizialmente presente, appare ora naturale che in certe classi di sistemi dina-

mici stocastici si introduca un livello di astrazione ancora piú alto e si parli di simboli e informazione. Piú di ogni altra caratteristica individuale, presentata e analizzata nel corso di questi quattro capitoli, si deve considerare questa possibilità di carattere generale come l'impronta caratteristica della complessità.

9. *Ancora una volta, la complessità algoritmica.*

A questo punto vale la pena di considerare nuovamente la connessione fra complessità fisico-chimica e il concetto di *complessità algoritmica* introdotto nel capitolo I, § 5. Come abbiamo visto, la complessità algoritmica è la lunghezza della descrizione piú corta di una data sequenza (finita). In questo senso, una sequenza completamente casuale generata da un processo di rumore possiede la massima complessità algoritmica possibile, essenzialmente uguale alla sua lunghezza, mentre una sequenza come XXX... possiede la minima complessità algoritmica possibile, poiché una singola istruzione è sufficiente a riprodurla completamente. Similmente, l'informazione si considera essere massima in una specifica sequenza casuale poiché la realizzazione di una particolare sequenza tratta dall'enorme numero di sequenze casuali equivale a localizzare il sistema molto nettamente nello spazio degli stati. L'informazione è zero nella sequenza di minima complessità poiché il risultato consiste unicamente in sequenze che sono tutte uguali.

Abbiamo già insistito sul fatto che la complessità degli oggetti naturali risiede da qualche parte fra questi due estremi, poiché oltre alla casualità essa implica anche qualche regolarità su larga scala. Il celebre aforisma di Erwin Schrödinger, che il DNA è un «cristallo aperiodico», racchiude questa dualità in modo particolarmente lucido e preciso.

Ora, gli stati autoorganizzati della materia permessi dalla fisica del non-equilibrio ci forniscono modelli proprio di questo tipo di complessità. Fra questi stati, il piú importante per i nostri scopi è la dinamica caotica. Infatti, l'instabilità del moto associato al caos permette al sistema di esplorare continuamente il suo spazio delle fasi, creando cosí informazione e complessità. D'altra parte, essendo il risultato di un sistema fisico, questi stati vengono prodotti con probabilità uno: il problema della selezione di una particolare sequenza da un gran numero di sequenze a priori ugualmente probabili, semplicemente non si pone. In un certo senso, il sistema dinamico che genera il caos agisce come un ef-

ficiente selettore che rigetta la maggior parte delle sequenze casuali e tiene solamente quelle compatibili con le sottostanti leggi di riproducibilità. Forse ugualmente importante, l'irreversibilità incorporata in queste leggi permette l'esistenza degli attrattori che hanno stabilità asintotica e dunque riproducibilità. Nel concludere questo capitolo è dunque legittimo affermare che abbiamo finalmente identificato un'origine fisica e qualche plausibile archetipo matematico della complessità.

Capitolo quinto

Verso una formulazione unificata della complessità

Le basi dinamiche della complessità costituiscono uno dei temi principali di questo libro. Per questo abbiamo bisogno sia della teoria dei sistemi dinamici sia dei concetti della teoria della probabilità, specialmente le catene di Markov. Nel capitolo II abbiamo sottolineato la distinzione fondamentale tra sistemi conservativi e sistemi dissipativi. La varietà di strumenti necessari per comprendere le basi dinamiche della complessità riflette l'ampiezza dei meccanismi che portano alla nascita della complessità, come il determinarsi di comportamenti caotici nei sistemi conservativi non integrabili o della rottura di simmetria nei sistemi dissipativi, o l'apparire di correlazioni a lungo raggio nelle condizioni di non-equilibrio.

Come sono correlati tra loro questi strumenti? Come possiamo passare dai sistemi conservativi a quelli dissipativi? Quale è il ruolo della probabilità? Questi sono i problemi che vogliamo trattare brevemente in questo capitolo. Poiché la dissipatività è fortemente correlata alla seconda legge della termodinamica e l'entropia stessa alla probabilità (cfr. capitolo IV, § 3), noi ritorniamo su uno dei problemi fondamentali della fisica: il significato e il ruolo dell'irreversibilità.

Questo problema ha turbato i fisici fin dalla formulazione della seconda legge della termodinamica nella metà del XIX secolo. La seconda legge fu il risultato di secoli di osservazioni, soprattutto nel campo dell'ingegneria e della fisica chimica, eppure è rimasta in qualche modo fuori dalle principali correnti della fisica. Essa introduce una quantità fisica, l'entropia, che dà al tempo una direzione privilegiata, la «freccia del tempo», per usare l'espressione di Arthur Eddington. Al contrario, fin dai tempi di Newton, la fisica ha posto come suo compito il raggiungimento di un livello fondamentale della realtà che è senza tempo. Le grandi rivoluzioni concettuali della relatività e della meccanica quantistica non hanno alterato sostanzialmente questa visione della fisica classica.

Di fronte a tale flagrante contraddizione tra le leggi dei sistemi dinamici conservativi e la seconda legge della termodinamica la maggioranza dei fisici ha considerato la seconda legge come generata da qualche forma di approssimazione sovrapposta alle leggi della dinamica. Max Born, per esempio, ha detto che «l'irreversibilità è una conseguenza dell'introduzione esplicita della nostra ignoranza nelle leggi fondamentali (della dinamica)».

Comunque dovrebbe essere chiaro dai molti esempi qui esaminati che l'irreversibilità gioca un ruolo fondamentale nella natura. Non possiamo semplicemente capire la maggior parte dei fenomeni attorno a noi, siano essi il formarsi di un cristallo, o la regolazione dei sistemi biologici, o la storia climatica, senza rivolgerci ai sistemi dissipativi. In campi molto differenti della fisica il ruolo attribuito all'irreversibilità diviene sempre maggiore. Chi si sarebbe aspettato che le particelle elementari sono in gran parte, se non tutte, instabili; e che processi irreversibili sembrano aver giocato un ruolo decisivo nei primi stadi dell'evoluzione del cosmo? Diventa pertanto impossibile rigettare l'irreversibilità in quanto contraria alla tradizionale e accettata visione della fisica. Prima di procedere alla formulazione della seconda legge come un principio dinamico, ci sembra appropriato riassumere alcune delle proprietà dei sistemi dinamici conservativi e dissipativi discussi nei precedenti capitoli.

1. *Proprietà generali dei sistemi dinamici conservativi.*

L'evoluzione di un sistema dinamico conservativo corrisponde a una trasformazione puntuale (che denoteremo con S_t) in un appropriato spazio delle fasi Γ:

$$\omega_t = S_t \omega_0 \qquad (5.1a)$$

dove ω_t e ω_0 sono punti dello spazio delle fasi al tempo t e 0 rispettivamente. Una fondamentale proprietà di questi sistemi, derivante dal teorema di Liouville, è che il flusso nello spazio delle fasi è quello di un fluido incomprimibile. In altre parole il volume si conserva nello spazio delle fasi (cfr. capitolo III, § 2). La dinamica può essere espressa in termini di un operatore unitario U_t che agisce sulla funzione di distribuzione (cfr. capitolo IV, § 3)

$$\rho_t(\omega) = U_t \rho_0(\omega) \qquad (5.1b)$$

$\rho_t(\omega)$ è la funzione di distribuzione al tempo t, e $\rho_0(\omega)$ è il suo valore iniziale[1]. Come abbiamo notato (cfr. equazione 4.22*b*), U_t forma un gruppo dinamico

$$U_t U_s = U_{t+s} (t, s \text{ positivo o negativo}) \tag{5.2}$$

Questa descrizione formale dei sistemi dinamici include classi di sistemi aventi comportamenti enormemente differenti. Nella fisica classica i sistemi integrabili come il pendolo senza attrito o il problema dei due corpi erano il centro dell'interesse. Sappiamo ora che questa è una classe eccezionalmente semplice che offre una idea inadeguata della vera complessità dei sistemi dinamici. La scoperta della varietà di comportamenti dinamici dei sistemi non integrabili segna uno dei maggiori passi avanti della fisica e della matematica contemporanee (cfr. capitolo III, § 8).

Quali sarebbero le proprietà di un mondo dominato dai sistemi dinamici conservativi?[2]. Forse il punto fondamentale, che abbiamo già ripetutamente menzionato, è che questo mondo sarebbe instabile, poiché la proprietà della stabilità asintotica, che permette a un sistema di dimenticare perturbazioni accidentali, non si applica ai sistemi dinamici conservativi (cfr. capitolo II, § 6).

Al piú si avrebbe la «stabilità orbitale» attorno ai punti ellittici (cfr. capitolo III, § 3), uno stato in cui le perturbazioni spostano il sistema da un'orbita all'altra. La situazione sarebbe peggiore attorno ai punti iperbolici, poiché qui le perturbazioni porterebbero all'instabilità. Inoltre vi è in generale abbondanza di punti ellittici e iperbolici. Come risultato dobbiamo aspettarci una estrema sensibilità alle condizioni iniziali. Abbiamo già visto un esempio di questo comportamento nel ferro di cavallo e presenteremo nel § 5, come illustrazione ancora piú semplice della casualità orbitale, la trasformazione del fornaio.

Il moto che coinvolge sfere rigide fornisce un altro esempio di sistema instabile. Un modello molto studiato è il modello di Lorentz, dal nome del fisico olandese Hendrick Antoon Lorentz. In questo modello si considera una piccola sfera che rimbalza sopra una serie di grandi sfere distribuite casualmente (cfr. figura 5.1). La traiettoria della sfera picco-

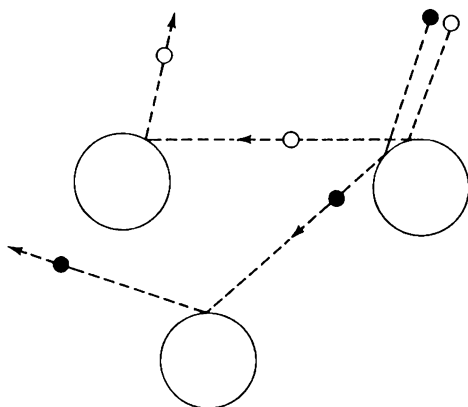
[1] Poiché il moto può essere espresso in termini di traiettorie descritte dall'equazione (5.1*a*), l'equazione (5.1*b*) è equivalente a $\rho_t(\omega_t) = \rho_0(\varsigma_0)$ o $\rho_t(\varsigma) = \rho_0(S - t_\varsigma)$.

[2] Nel senso che i macroscopici osservabili seguirebbero una dinamica conservativa, temporalmente reversibile.

la è ben definita. Ogni volta che introduciamo la piú piccola incertezza nelle condizioni iniziali, essa viene amplificata attraverso le collisioni successive. Si noti che le reazioni chimiche possono spesso essere considerate come il risultato di collisioni tra sfere rigide nelle quali l'energia cinetica relativa (la cosí detta energia di attivazione) va oltre una certa soglia. Nuovamente ci si aspetta che questi moti siano fortemente instabili.

Mentre il volume (la misura) dello spazio delle fasi è sempre conservato durante l'evoluzione, la sua forma, come risultato dell'instabilità, viene fortemente distorta. Come vedremo nel § 5, il volume iniziale può frammentarsi in pezzi sempre piú piccoli col passare del tempo. È questa deformazione, o frammentazione, a fornire l'apparenza di un avvicinamento all'equilibrio in cui tutti i punti saranno uniformemente distribuiti nella regione accessibile dello spazio delle fasi. Dopo Newton, per quasi tre secoli, la dinamica classica è apparsa come una scienza chiusa, che ci metteva in grado di calcolare ogni evoluzione a partire da principî primi e da ben definiti dati iniziali. Oggi vediamo che ciò si applica solo a classi limitate di sistemi dinamici: per sistemi sufficientemente instabili ogni regione dello spazio delle fasi, quale che sia la sua dimensione, contiene traiettorie divergenti. Per poter parlare di una singola traiettoria ben definita avremmo bisogno di dati iniziali aventi

Figura 5.1

Rappresentazione schematica dell'instabilità della traiettoria di una piccola sfera che rimbalza su sfere piú grandi. La piú piccola imprecisione riguardo la posizione iniziale della piccola sfera rende impossibile prevedere quale sfera grande colpirà dopo alcune collisioni.

una precisione assolutamente infinita. In altre parole avremmo bisogno di una quantità infinita di informazione (rappresentata dall'infinito numero di cifre necessarie per descrivere i dati iniziali). È l'eliminazione di questa informazione infinita che porta, come vedremo, all'irreversibilità[3].

Nel mondo dei sistemi dinamici instabili possiamo guardare al mondo esterno solo attraverso una «finestra». Non possiamo quindi che prendere atto del crollo dell'ideale della conoscenza totale che per tre secoli ha ossessionato la scienza occidentale.

2. *Proprietà generali dei sistemi dissipativi.*

I sistemi dinamici dissipativi presentano una straordinaria varietà. Abbiamo visto molti esempi riferiti alla chimica, alla fluidodinamica o delle scienze naturali. Sappiamo che, contrariamente ai sistemi conservativi, i sistemi dissipativi non conservano in generale nessuna misura nello spazio delle fasi (cfr. capitolo III, § 2). Inoltre, essi hanno in comune l'elemento cruciale di non essere invarianti rispetto all'inversione temporale. Ne segue che asintoticamente, quando $t \to +\infty$, si raggiungerà una qualche situazione privilegiata. Se non imponiamo nessun vincolo questa sarà la situazione di equilibrio. Ad esempio, la concentrazione o la temperatura descritta dalle equazioni 2.7*b* e 2.7*c* tenderà all'uniformità; analogamente la concentrazione corrispondente a un miscuglio reattivo tenderà alla legge di azione di massa, che corrisponde nuovamente all'equilibrio termodinamico. Se si impongono i vincoli di non-equilibrio l'evoluzione tenderà a qualche insieme attrattore, sia esso un punto attrattore, un ciclo limite, o uno «strano» attrattore, come nel capitolo III, § 10. È essenziale che in tutti i sistemi dissipativi vi sia una direzione privilegiata del tempo. Potremmo immaginare un mondo in cui alcuni sistemi biologici invecchiano, mentre altri diventano piú giovani; un mondo in cui alcuni sistemi dissipativi tendono all'equilibrio per $t \to +\infty$ mentre altri fanno cosí per $t \to -\infty$. Ma questo non è ovviamente il nostro mondo, in cui è presente una universale rottura di simmetria temporale. Fortunatamente è cosí, come notato da Norbert Wiener: in qualsiasi mondo entro il quale possiamo comunicare, la direzione del tempo deve essere uniforme.

[3] Ci limitiamo qui alla meccanica classica. Nella teoria quantistica il decadimento di particelle instabili è una fonte aggiuntiva di irreversibilità.

Contrariamente alla casualità orbitale dei sistemi dinamici conservativi, troviamo qui la stabilità asintotica. Questo rende possibile la comparsa di processi in delicato equilibrio, come ad esempio vedremo in biologia. La casualità può tuttavia ancora sussistere a livello degli attrattori.

Nel capitolo III, § 10, abbiamo studiato diversi modelli in cui è possibile osservare un comportamento caotico. Tuttavia gli esponenti di Ljapunov che, secondo l'*Appendice* A, caratterizzano l'instabilità del moto, vengono ora determinati da quantità macroscopiche come le velocità chimiche e i coefficienti di diffusione e non dalla dinamica microscopica.

Non solo il comportamento dissipativo è associato a una universale rottura di simmetria; possiamo in aggiunta associare ai processi dissipativi una funzione di Ljapunov come l'entropia (o la quantità H) che cambia in modo monotòno nel tempo. Infatti tutti i processi dissipativi che abbiamo studiato portano a una produzione di entropia positiva. La stessa esistenza di sistemi dinamici dissipativi è una manifestazione della seconda legge della termodinamica.

3. *La ricerca dell'unificazione.*

C'è ancora un altro punto di vista che ci permette di contrapporre i sistemi dinamici conservativi ai sistemi dissipativi. Sembra impossibile trovare un livello di descrizione piú fondamentale a partire dal quale i sistemi dinamici conservativi emergano come il risultato di approssimazioni. Naturalmente, se introduciamo la teoria quantistica, le traiettorie classiche diventano le medie sui pacchetti d'onda. Questo non porta tuttavia a un nuovo punto di vista, poiché l'equazione di Schrödinger – l'equazione fondamentale della teoria quantistica – è deterministica e temporalmente reversibile, proprio come le equazioni della dinamica classica[4].

La situazione è del tutto differente per i sistemi dinamici dissipativi. Consideriamo ad esempio la cinetica chimica. Le equazioni incontrate qui non corrispondono a un «livello fondamentale», poiché, come abbiamo visto, esse risultano da approssimazioni in cui le *fluttuazioni*

[4] Analogamente, non servirebbe sostituire la dinamica delle particelle con equazioni di campo, come quelle impiegate nell'elettrodinamica o nella relatività. Anche le fondamentali equazioni di campo sono deterministiche e temporalmente reversibili.

vengono trascurate (cfr. specialmente l'equazione 4.15). Osserviamo anche *correlazioni* che riflettono il comportamento coerente che nasce nelle situazioni di non-equilibrio. Il livello fondamentale dei sistemi dissipativi deve dunque essere considerato in congiunzione con un *approccio probabilistico* come quello fornito dalla teoria delle catene di Markov studiate nel capitolo IV.

Può questa formulazione markoviana essere estesa dallo spazio dei macroscopici osservabili (a cui si limitava il capitolo IV) all'intero spazio delle fasi Γ, per il quale fino ad ora erano disponibili solo descrizioni reversibili (cfr. le equazioni 5.1)? Possiamo tentare introducendo in Γ una funzione di distribuzione $\tilde{\rho}(\omega)$, che, in contrasto con la meccanica classica, evolve ora secondo un processo di Markov. In analogia con la 5.1 possiamo scrivere per $\tilde{\rho}_t$ al tempo t

$$\tilde{\rho}_t(\omega) = W_t\tilde{\rho}_0(\omega) \tag{5.3}$$

si intende che $\tilde{\rho}_t(\omega)$ è la soluzione di una equazione principale che descrive il processo di Markov. Come abbiamo visto la descrizione reversibile basata sull'equazione di Liouville soddisfa la proprietà di gruppo 5.2. Al contrario, W_t soddisfa la condizione di semigruppo.

$$W_t W_s = W_{t+s} \qquad (t, s \geqslant 0) \tag{5.4}$$

Il processo di Markov 5.3, con la condizione 5.4, porta il sistema all'equilibrio, o a qualche stato di non-equilibrio caratterizzato da opportuni attrattori quando vengono introdotti vincoli. Riassumendo, nella dinamica conservativa sembra non esservi posto per processi evolutivi probabilistici genuini; tali processi sono invece fondamentali nella descrizione dei sistemi dissipativi. Dunque l'elucidazione dei rapporti tra sistemi dinamici conservativi e dissipativi implica una chiarificazione dei rapporti tra dinamica deterministica e probabilità: una volta chiariti questi anche il rapporto tra dinamica e seconda legge sarà evidente. Poiché le leggi fondamentali della dinamica conservativa reversibile possono essere espresse dall'equazione 5.1, mentre i processi probabilistici genuini soddisfano le relazioni delle equazioni 5.3 e 5.4, dobbiamo discutere il modo in cui tutte queste espressioni possono essere collegate.

4. *Probabilità e dinamica.*

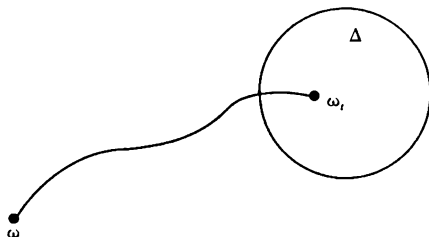
La quantità fondamentale nella teoria dei processi di Markov è la probabilità di transizione (cfr. capitolo IV, § 2). Poiché vogliamo stabilire una descrizione probabilistica genuina dello spazio delle fasi Γ, consideriamo la probabilità di transizione $P_t(\omega|\Delta)$ dal punto ω a qualche dominio Δ durante il tempo t. Questa quantità deve essere un numero positivo fra zero e uno. Qui arriviamo immediatamente a una fondamentale differenza tra la descrizione markoviana e la descrizione in termini di traiettorie. Supponiamo che una traiettoria vada da ω a ω_t, allora ovviamente abbiamo

$$P_t(\omega|\Delta) = 1 \quad \text{se } \omega_t \text{ è in } \Delta \tag{5.5}$$
$$= 0 \quad \text{se } \omega_t \text{ non è in } \Delta$$

Al contrario, in un processo di Markov almeno alcune delle probabilità di transizione non sono né zero né uno.

Non abbiamo nessuna ragione di ritenere che la descrizione dinamica dell'equazione 5.1 sia errata. Potrebbe tuttavia essere incompleta, dal momento che non incorpora l'elemento di non-località che ovviamente è necessario per superare il divario tra sistemi conservativi e dissipativi, portando dalle probabilità di transizione degenerate (equazione 5.5) a quelle markoviane corrette. In vista di questa situazione si è tentati di correlare le descrizioni delle equazioni 5.1 e 5.3 con la trasformazione

$$\tilde{\rho} = \Lambda\rho \tag{5.6}$$

Figura 5.2
Probabilità di transizione $P_t(\omega|\Delta) = 1$, caratteristica di una descrizione deterministica basata sulle traiettorie.

dove l'operatore di trasformazione Λ agente su ρ dovrebbe incorporare la non-località mancante e dipende naturalmente dalle leggi della dinamica. Una volta riusciti a determinare Λ, l'equazione 5.6 fornirà un collegamento tra la funzione di distribuzione ρ, che evolve secondo la dinamica reversibile, e la funzione di distribuzione $\tilde{\rho}$, che evolve invece secondo un processo di Markov.

La relazione in 5.6 implica un notevole «intrecciamento» fra U_t e W_t. Infatti, inserendo le equazioni 5.1*b* e 5.3 nella 5.6 otteniamo

$$\Lambda U_t \rho_0 = W_t \Lambda \rho_0 \tag{5.7}$$

Poiché questo è vero per qualsiasi ρ_0 concludiamo

$$\Lambda U_t = W_t \Lambda \tag{5.8}$$

Se Λ ammette l'inverso, la 5.8 si può riscrivere:

$$\begin{aligned} W_t &= \Lambda U_t \Lambda^{-1} \\ U_t &= \Lambda^{-1} W_t \Lambda \end{aligned} \tag{5.9}$$

In questo quadro, la teoria della probabilità è un'«immagine» della dinamica mediata dalla trasformazione Λ. Inversamente la dinamica sottostante a un processo di Markov si può ritrovare ogni qualvolta siamo in grado di identificare Λ. Il problema centrale che affrontiamo nel creare un ponte fra la dinamica e la probabilità risiede dunque nella costruzione di Λ. Vediamo ora una classe di sistemi dinamici conservativi per i quali questa costruzione può essere chiaramente realizzata.

5. *La trasformazione del fornaio.*

Consideriamo la classe di sistemi dinamici altamente instabili chiamati flussi K (K sta per Kolmogorov). Nel capitolo III, § 9, abbiamo studiato l'esempio del ferro di cavallo. Qui esamineremo un altro celebre modello, la trasformazione del fornaio, schematizzata nella figura 5.3. Nello spazio delle fasi $0 \leqslant x \leqslant 1$, $0 \leqslant y \leqslant 1$ (il quadrato unitario), definiamo la mappa

$$(5.10)\qquad \begin{aligned} x' &= 2x, & y' &= \frac{y}{2} & &\text{se } 0 \leqslant x < \frac{1}{2} \\ x' &= 2x - 1, & y' &= \frac{y+1}{2} & &\text{se } \frac{1}{2} \leqslant x \leqslant 1 \end{aligned}$$

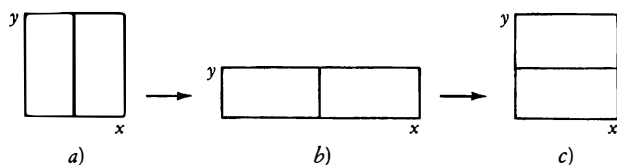
Questa trasformazione corrisponde a un processo di evoluzione discreto che conserva la misura e che si verifica secondo tempi regolari, ad esempio a ogni secondo (qui la misura è l'area della superficie). La trasformazione definisce perciò un sistema dinamico conservativo. Esso è anche ovviamente invertibile, poiché il quadrato unitario è definito sulla mappa di se stesso (al contrario di ciò che accade nella trasformazione del ferro di cavallo).

La reversibilità temporale della dinamica è in qualche modo piú sottile da stabilire. Risolviamo le equazioni 5.10 per x e y:

$$(5.11)\qquad \begin{aligned} x &= \frac{x'}{2}, & y &= 2y' & &0 \leqslant y < \frac{1}{2} \\ x &= \frac{x'+1}{2}, & y &= 2y' - 1 & &\frac{1}{2} \leqslant y \leqslant 1 \end{aligned}$$

Vediamo che le equazioni 5.11 sono identiche alla versione temporalmente invertita delle equazioni 5.10, a patto che x e y siano scambiate. Questo scambio è analogo all'inversione dei momenti (mantenendo immodificate le coordinate spaziali) quando si inverte il tempo nelle equazioni della meccanica hamiltoniana (cfr. cap II, § 1).

Nelle equazioni 5.10 osserviamo che una direzione, x, si espande e che l'altra, y, si contrae. Dopo n trasformazioni consecutive la distanza tra due punti sull'asse delle x sarà moltiplicata secondo un fattore

Figura 5.3
La trasformazione del fornaio. Il quadrato unitario, *a*), è portato su un rettangolo di lati $\frac{1}{2} \times 2$ unità, *b*). La metà destra del rettangolo è quindi sovrapposta alla parte sinistra a formare un nuovo quadrato *c*).

$2^n = e^{n \ln 2}$. Secondo le definizioni date nell'*Appendice* A avremo quindi un esponente di Ljapunov positivo,

$$\sigma_1 = \ln 2 \tag{5.12}$$

Questo stabilisce il carattere caotico della dinamica. Naturalmente, poiché il sistema è conservativo, il secondo esponente di Ljapunov è negativo, $\sigma_2 = -\ln 2$ (cfr. equazione A1.27).

Si può anche osservare, attraverso la ripetizione del processo indicato nella figura 5.3, che col passare del tempo, ogni regione definita viene frammentata in strisce sempre piú piccole, finché non diviene in pratica uniformemente distribuita sul quadrato unitario. Questo processo è illustrato nella figura 5.4.

Una buona comprensione del meccanismo della trasformazione del fornaio si ottiene rappresentandola come un particolare tipo di processo noto come *traslazione di Bernoulli*. Per fare questo possiamo associare a ogni punto (x, y) del quadrato unitario una sequenza doppiamente infinita di numeri $\{u_n\}$ definita dalla rappresentazione binaria:

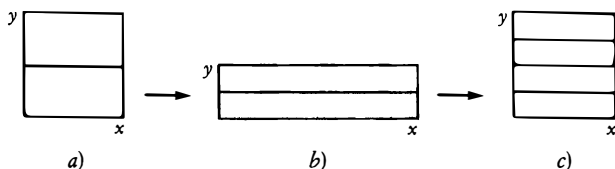
$$x = \sum_{n=-\infty}^{0} \frac{u_n}{2^{-n+1}}, \qquad y = \sum_{n=1}^{\infty} \frac{u_n}{2^n} \tag{5.13}$$

dove ogni u_n assume valori 0 o 1. Inserendo questa rappresentazione nelle equazioni 5.10, otteniamo:

$$\frac{u_0'}{2} + \frac{u'_{-1}}{2^2} + \cdots = x' = \frac{u_{-1}}{2} + \frac{u_{-2}}{2^2} + \cdots$$

$$\frac{u_1'}{2} + \frac{u_2'}{2^2} + \cdots = y' = \frac{u_0}{2} + \frac{u_1}{2^2} + \cdots$$

Figura 5.4

Iterazioni successive della trasformazione del fornaio portano alla frammentazione sempre maggiore delle aree bianche e grigie.

L'evoluzione temporale indotta dalla trasformazione del fornaio corrisponde perciò alla traslazione

$$u_n' = u_{n-1} \tag{5.14}$$

a cui ci siamo riferiti come traslazione di Bernoulli. Vediamo quindi che, come ci si aspettava, l'informazione contenuta nelle condizioni iniziali include tutta la storia passata e futura del sistema.

Si noti che il valore u_0 determina se il punto rappresentativo dello spazio delle fasi è nella metà sinistra ($u_0 = 0$) o nella metà destra ($u_0 = 1$) del quadrato unitario. Poiché i valori u_0, u_{-1}, u_{-2}, ... possono essere determinati con il lancio di una moneta, le iterazioni temporali di u_0, $u_0' = u_{-1}$, $u_0'' = u_{-2}$ avranno le stesse proprietà casuali del lancio della moneta. Questo mostra come l'evento del sistema di essere a sinistra o a destra può essere considerato come un processo di Bernoulli, uno dei piú familiari esempi di processi casuali della teoria delle probabilità.

La trasformazione del fornaio possiede anche l'importante proprietà della *ricorrenza* degli stati. Consideriamo un punto (x, y) per il quale la sequenza $\{u_n\}$ nella rappresentazione binaria (equazioni 5.13), sia infinita o infinitamente periodica. In questo caso x e y sono razionali. Poiché tutti gli u_n sono traslati nello stesso modo dalla dinamica, tutti gli stati di questo tipo ripeteranno esattamente il ciclo dopo un certo periodo di tempo. Ancora piú importante è che lo stesso si mostra vero, in un senso piú debole attualmente da specificarsi, per quasi tutti gli altri stati (eccetto che per un insieme di misura zero). Per vedere questo consideriamo la rappresentazione binaria di un punto irrazionale (x, y), che contiene un'infinità di cifre non banali che non si ripetono. Si può mostrare che quasi tutti gli irrazionali (eccetto che per un insieme di misura zero) contengono ogni sequenza finita di cifre infinitamente spesso. Cosí una data sequenza di, diciamo, $2m$ cifre attorno alla posizione 0, che determina lo stato fino a un errore di 2^{-m}, riapparirà infinitamente spesso sotto l'effetto della traslazione. Poiché m può essere scelto grande a piacere (sebbene finito), ciò significa che quasi ogni stato ritornerà un infinito numero di volte arbitrariamente vicino a ogni punto, inclusa naturalmente la posizione iniziale. In altri termini la maggior parte delle traiettorie sarà *ergodica* e traverserà l'intero spazio delle fasi. Questo è il famoso teorema della ricorrenza di Poincaré, originariamente dimostrato per una classe molto generale di sistemi dinamici conservativi. Per molto tempo la ricorrenza, assieme alla reversibilità temporale, è stata avanzata come argomento essenziale contro l'esistenza di processi

genuinamente dissipativi. Questa concezione, come vedremo tra breve, non può piú essere sostenuta.

Riassumendo, la dinamica descritta dalla trasformazione del fornaio è conservativa, invertibile, temporalmente reversibile, ricorrente e caotica. La manifestazione di tutte queste proprietà fa della trasformazione del fornaio un esempio particolarmente utile, poiché queste stesse proprietà caratterizzano i sistemi dinamici del mondo reale che mostrano un comportamento complesso. Colpisce il fatto che, nonostante la reversibilità e la ricorrenza, la proprietà della caoticità ci consentirà di stabilire la genuina irreversibilità con la costruzione, senza introdurre approssimazioni di sorta, di una descrizione piú appropriata.

La dinamica dei sistemi conservativi comprende le leggi del moto e le condizioni iniziali. Qui le leggi del moto sono semplici; analizziamo invece il concetto di dati iniziali piú in dettaglio. I dati iniziali di una singola traiettoria corrispondono a un insieme infinito $\{u_n\}$ ($n = -\infty$ a $+\infty$). Ma nel mondo reale guardiamo soltanto attraverso una finestra finita. Nel caso in oggetto questo significa che controlliamo un numero arbitrario, ma limitato, di valori di u_n. Supponiamo che tale finestra corrisponda a

$$u_{-3}u_{-2}u_{1}u_{-0} \cdot u_{1}u_{2}u_{3} \tag{5.15}$$

tutti gli altri valori essendo sconosciuti (il punto indica la separazione tra x e y). La traslazione di Bernoulli prevede che al passo successivo l'equazione 5.15 sia sostituita dalla

$$u_{-4}u_{-3}u_{-2}u_{-1} \cdot u_{0}u_{1}u_{2} \tag{5.16}$$

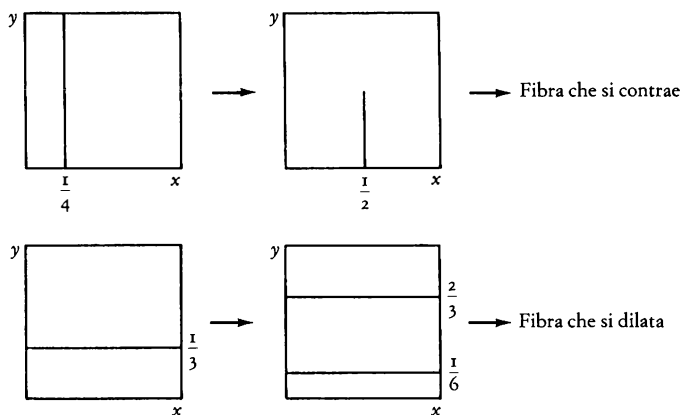
che contiene il valore non noto u_{-4}. Piú precisamente, a causa dell'esistenza di un esponente di Ljapunov positivo, abbiamo la necessità di conoscere la posizione iniziale con una accuratezza di $N + n$ cifre per poter determinare la posizione con una accuratezza di N cifre dopo ogni iterazione.

Nella meccanica statistica il modo tradizionale di affrontare questa situazione sarebbe stato di introdurre una grossolana distribuzione di probabilità nel quadrato unitario. Tale distribuzione non è definita su un singolo punto, ma su regioni, come originariamente proposto da Paul e Tatiana Ehrenfest. Tuttavia, due punti in un oggetto in espansione, anche se non distinguibili con misurazioni di una data precisione finita al tempo 0, saranno separati piú tardi e potranno allora essere di-

stinti. La grossolanità tradizionale perciò non tiene il passo con l'evoluzione dinamica. Per questo avremo bisogno di un metodo piú sofisticato, che svilupperemo nei due paragrafi che seguono.

6. *Varietà aventi rottura di simmetria temporale.*

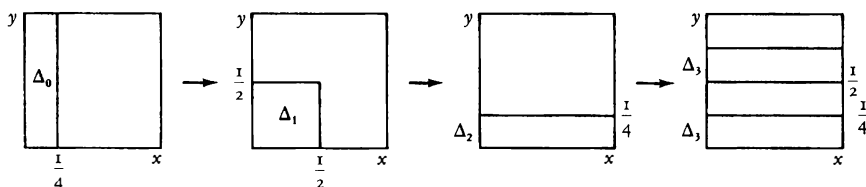
A dispetto del fatto che i flussi K sono temporalmente reversibili, essi contengono oggetti aventi rottura di simmetria temporale. Questi oggetti sono di dimensioni inferiori rispetto all'intero spazio delle fasi, e nel caso della trasformazione del fornaio sono «fibre che si contraggono» e «fibre che si dilatano», come mostrato dalla figura 5.5. Le fibre che si contraggono sono linee verticali che si contraggono secondo un fattore 2 a ogni trasformazione del fornaio; le fibre che si dilatano sono linee orizzontali che si espandono secondo lo stesso fattore. Ogni punto dello spazio delle fasi è l'intersezione di una fibra che si contrae e di una che si dilata. Varietà che si contraggono e si dilatano sono gli inversi temporali gli uni degli altri, e sono ovviamente *asimmetrici rispetto al tempo*. Tutti i punti di una varietà che si contrae tendono verso lo stesso futuro. All'opposto, i punti su una fibra che si dilata copriranno densamente lo spazio delle fasi per $t \to +\infty$. In linguaggio probabilistico ciò corrisponde all'avvicinarsi alla distribuzione uniforme d'equilibrio nello spazio delle fasi.

Figura 5.5
Fibre che si contraggono e che si dilatano nella trasformazione del fornaio.

Quale che possa essere la nostra precisione nel determinare uno stato, ci sarà sempre una distanza ε al di sotto della quale non possiamo distinguere i punti nelle fibre che si contraggono. Poiché i punti su queste fibre appartengono a traiettorie convergenti, *tutte le traiettorie che iniziano entro* ε *devono essere considerate indiscernibili nel futuro*. Ciò rende immediatamente irrilevante il teorema di Liouville che esprime la conservazione della misura nello spazio delle fasi. Una semplice discussione è la seguente. Consideriamo un sistema dinamico e supponiamo che la funzione di distribuzione ρ sia confinata, per $t = 0$ in un'area Δ_0. Col passare del tempo la dimensione della dilatazione crescerà esponenzialmente, mentre la dimensione della contrazione si restringerà a ε *e andrà considerata costante di qui in poi* (la procedura è illustrata dalla figura 5.6). Come risultato si avrà che la misura Δ crescerà da quel momento in poi fino a raggiungere la misura dello spazio delle fasi totale. Si noti che questa conclusione non risulta da una approssimazione. È invece la conseguenza del fatto fondamentale che le misurazioni (o, piú in generale, gli effetti delle interazioni fisiche) si riferiscono a *regioni* di misura finita nello spazio delle fasi, non a punti matematici. Naturalmente si è sempre compreso che condizioni iniziali esatte dovessero ritenersi una idealizzazione. La novità sta nel fatto che l'abbandono di questa idealizzazione porta a conseguenze di grande importanza in sistemi altamente instabili come i flussi K.

7. *La trasformazione di rottura di simmetria* Λ.

Indichiamo ora come le idee avanzate nel paragrafo precedente conducano alla costruzione della trasformazione di rottura di simmetria Λ, che collega processi dinamici e processi probabilistici.

Figura 5.6

Inadeguatezza del teorema di Liouville, quando l'osservazione comprende solo una finestra finita (qui $\varepsilon = \frac{1}{4}$), sulle fibre che si contraggono. La misura del volume iniziale Δ_0 resta invariante per alcune interazioni, ma alla fine cresce fino a raggiungere la misura dello spazio delle fasi totale.

A questo scopo è utile rappresentare lo stato del sistema dinamico instabile considerato in termini di oggetti, a cui ci riferiremo come *partizioni*, $\{\chi_n\}$. Nel caso della trasformazione del fornaio costruire tali partizioni è molto facile. Infatti, sia χ_0 la funzione che assume i valori -1 nella metà sinistra del quadrato unitario, e $+1$ nella destra. Definiamo χ_n come il risultato di n successive trasformazioni (per n intero positivo o negativo) operate su χ_0:

$$\chi_n = U^n \chi_0 \tag{5.17}$$

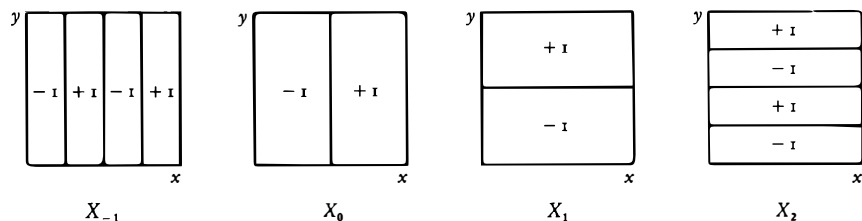
dove U denota l'operatore di evoluzione per le funzioni di distribuzione (equazione 5.1*b*). Alcune di queste partizioni sono rappresentate nella figura 5.7. Si noti che χ_0 è interamente determinato dal valore u_0 nella rappresentazione dell'equazione 5.13, mentre nella 5.17 è determinato dal valore u_n.

Vediamo ora come possiamo descrivere la dinamica in termini di partizioni. Per fare questo è utile caratterizzare le partizioni χ_n con la loro «età» n. Questo concetto, che alla luce dell'equazione 5.17 è del tutto naturale, può essere formalizzato ed esteso per mezzo di un importante risultato di Baidyanath Misra, che stabilisce l'esistenza di un operatore T che ammette χ_n e n rispettivamente come autofunzioni e autovalori $T\chi_n = n\chi_n$. Effettivamente ogni autovalore n di questo operatore mostra una degenerazione infinita, dacché oltre a χ_n, T possiede tutti i prodotti

$$\Phi_n = \chi_{j_1}\chi_{j_2}\cdots\chi_n \qquad (j_1 < j_2 < \cdots < n)$$

come autofunzioni. Scriviamo perciò

$$T\Phi_n = n\Phi_n \qquad (-\infty \leqslant n \leqslant \infty) \tag{5.18a}$$

Figura 5.7
Rappresentazione della dinamica della trasformazione del fornaio in termini di partizioni.

Per completare la definizione di T sarà necessario che $\Phi \equiv 1$ sia una autofunzione aggiuntiva con autovalore 1

$$T \cdot 1 = 1 \tag{5.18b}$$

e che T sia un operatore hermitiano lineare. La funzione Φ_n assieme a $\Phi = 1$ può essere mostrata per formare un insieme completo.

Si noti che l'effetto di T su una partizione χ_n è differente da quello dell'operatore dinamico U, poiché secondo la 5.17

$$U\chi_n = \chi_{n+1} \tag{5.19}$$

Come conseguenza di questa differenza T e U sono operatori che non commutano. Dalle equazioni 5.18 e 5.19

$$TU\chi_n = T\chi_{n+1} = (n+1)\chi_{n+1}$$

D'altra parte

$$UT\chi_n = Un\chi_n = n\chi_{n+1}$$

Cosí,

$$(TU - UT)\chi_n = \chi_{n+1} \neq 0$$

È notevole che in questo modo la dinamica classica porta a un'algebra non commutativa, non diversa da quella della meccanica quantistica.

Introduciamo ora la condizione che la trasformazione di rottura di simmetria Λ sia un operatore lineare hermitiano che è una funzione decrescente non negativa di T, nel senso che

$$\begin{gathered} \Lambda(T)\Phi_n = \lambda_n \Phi_n \\ \Lambda(T)1 = 1 \end{gathered} \tag{5.20}$$

con

$$1 \geq \cdots \geq \lambda_n \geq \lambda_{n+1} \geq \cdots \geq 0$$

e

$$\lambda_n \to \begin{cases} 1 & \text{per } n \to -\infty \\ 0 & \text{per } n \to +\infty \end{cases}$$

Poiché la partizione Φ_n, assieme alla distribuzione uniforme 1, forma

un insieme completo, possiamo espandere ogni funzione di distribuzione ρ come

$$(5.21) \qquad \rho = 1 + \sum_{n=-\infty}^{+\infty} c_n \Phi_n$$

Scrivendo questa relazione abbiamo soppresso, per maggiore semplicità, un indice addizionale riferito alla degenerazione degli autovalori di T. Utilizzando le equazioni 5.20 e $\tilde{\rho} = \Lambda\rho$ segue che

$$(5.22) \qquad \tilde{\rho} = 1 + \sum_{n=-\infty}^{+\infty} c_n \lambda_n \Phi_n$$

Si noti che in un certo punto dello spazio delle fasi il passaggio da ρ a $\tilde{\rho}$ coinvolge tutti i punti dello spazio delle fasi. Questo determina la attesa non-località dell'operatore Λ.

Tutti gli stati Φ_n entrano in ρ come tali, essendo i loro effetti ponderati soltanto con i fattori c_n prescritti dalle condizioni iniziali. Sotto questo aspetto nessuna distinzione viene fatta tra passato ($n<0$) e futuro ($n>0$). Al contrario in $\tilde{\rho}$, per date condizioni iniziali, il contributo dei Φ_n riguardanti il futuro è smorzato ($\lambda_n \to 0$ per $n \to \infty$). Ora, ricordiamo che la partizione χ_n è determinata dal valore u_n. In aggiunta, il valore u_n con $n>0$ specifica interamente le fibre che si contraggono (cfr. le equazioni 5.13). Pertanto lo smorzamento $\lambda_n \to 0$ esprime il fatto che perdiamo informazione concernente la localizzazione su queste fibre. Ciò corrisponde esattamente all'idea dell'indiscernibilità dei punti tra loro vicini. Per esempio, con $\lambda_n = 1$ per $n \leqslant 2$ e $\lambda_n = 0$ per $n>2$, il valore naturale della distanza ε, introdotta nel § 6, è $\varepsilon = \frac{1}{8}$. Nel paragrafo citato abbiamo esposto intuitivamente la ragione per cui l'indiscernibilità rende inutile il teorema di Liouville.

Verifichiamo che per la nuova quantità $\tilde{\rho}$ possiamo ora definire una funzione che evolve in modo monotòno, in conformità con la seconda legge della termodinamica (cfr. capitolo IV, § 3). Consideriamo la cosiddetta funzione $\mathcal{H}$

$$(5.23) \qquad \mathcal{H}_{\tilde{\rho}} = \int \tilde{\rho}^2 \, d\omega$$

altre funzioni convesse, come $\int \tilde{\rho} \ln \tilde{\rho} \, d\omega$ possono essere introdotte con risultati simili. Supponiamo che al tempo $t = 0$ ρ e $\tilde{\rho}$ siano dati rispettivamente dalle equazioni 5.21 e 5.22; allora utilizzando la 5.19 e il fatto ovvio che $U \cdot 1 = 1$

$$\rho_1 = U\rho_0 = 1 + \sum_n c_n \chi_{n+1} \tag{5.24}$$

$$\tilde{\rho}_1 = \Lambda\rho_1 = 1 + \sum_n c_n \lambda_{n+1} \chi_{n+1} \tag{5.25}$$

Come risultato otteniamo dalla 5.23

$$\mathcal{H}_{\tilde{\rho}_1} - \mathcal{H}_{\tilde{\rho}_0} = \sum_n (\lambda_{n+1}^2 - \lambda_n^2) c_n^2 \leqslant 0 \tag{5.26}$$

ossia $\mathcal{H}_{\tilde{\rho}}$ decresce monotonamente col tempo. All'opposto la quantità $\mathcal{H}_\rho = \int \rho^2 \, d\omega$ espressa in termini dell'originale funzione di distribuzione ρ rimane costante. Il risultato ora ottenuto giustifica l'assunzione di monotonicità fatta introducendo l'operatore Λ (cfr. le equazioni 5.20). Di fatto la monotonicità dei λ_n è sia necessaria sia sufficiente per la validità dell'equazione 5.26.

Riassumendo, abbiamo raggiunto il nostro scopo di costruire la trasformazione Λ e di ottenere una formulazione microscopica della seconda legge della termodinamica. Questo evidenzia l'importanza della distribuzione $\tilde{\rho}$. S'intende che, di fatto, non solo l'entropia (tramite $\mathcal{H}_{\tilde{\rho}}$) ma anche tutti gli altri osservabili vanno espressi in termini di $\tilde{\rho}$.

Fin qui abbiamo solo ipotizzato che l'operatore di trasformazione $\Lambda(T)$ sia una funzione decrescente, ma altrimenti indeterminata, di T. Non possiamo diffonderci qui in una discussione dettagliata della forma di $\Lambda(T)$, né, in alternativa, di λ_n come una funzione di n. Poiché l'unico parametro intrinseco del nostro modello è l'esponente di Ljapunov (equazione 5.12), è ragionevole aspettarsi che almeno per n grande e positivo

$$\lambda_n \sim e^{-n/\tau} \tag{5.27}$$

dove $1/\tau = \ln 2$ è l'esponente di Ljapunov. Quale che sia la forma dettagliata di λ_n, le proprietà fondamentali di Λ sono la non-località e rottura di simmetria temporale. Queste sono le proprietà che portano dai sistemi dinamici conservativi instabili ai sistemi dissipativi.

8. *Insiemi di Gibbs e insiemi di Boltzmann.*

La dinamica classica porta a un punto di trasformazione. Nel linguaggio degli insiemi di Gibbs un punto nello spazio delle fasi corri-

sponde un insieme di Gibbs rappresentato da una funzione δ. Il moto $\omega_0 \rightarrow \omega_t$ trasforma una funzione δ un un'altra.

$$\delta_{\omega_t} = U_t \, \delta_{\omega_0} \qquad (5.28)$$

dove $\tilde{\delta}_{\omega_0} = \delta(\omega - \omega_0)$, ecc. La rappresentazione della funzione δ è pertanto invariante sotto l'azione dell'operatore unitario U_t. Applichiamo ora la trasformazione Δ a δ_{ω_0}

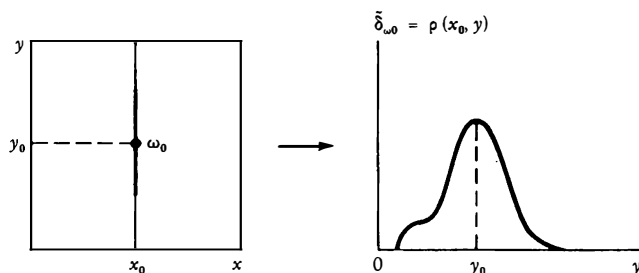
$$\tilde{\rho}_{\omega_0} = \Lambda \, \delta_{\omega_0} \qquad (5.29)$$

in generale $\tilde{\delta}_{\omega_0}$ non è strettamente localizzata. Tuttavia a causa della monotonicità dei λ_n è concentrata molto fortemente sulla fibra che si contrae con ω_0, lungo la quale presenta una distribuzione piú o meno larga con un massimo attorno a ω_0. La figura 5.8 mostra un esempio in cui $\tilde{\delta}_{\omega_0}$ è interamente concentrato sulla fibra che si contrae. L'evoluzione temporale di $\tilde{\delta}_{\omega_0}$ è data naturalmente dall'equazione 5.3, $\tilde{\delta}_{\omega_t} = W_t \, \tilde{\delta}_{\omega_0}$. Nel caso della figura si può mostrare che l'azione di W_t ammonta a uno spostamento rigido del supporto di $\tilde{\delta}$.

Nella rappresentazione tradizionale della dinamica le unità di base sono punti; ora sono le entità non locali $\tilde{\delta}_{\omega_0}$ che sono essenzialmente associate a segmenti delle fibre che si contraggono. Anziché di una singola traiettoria, rappresentata da funzioni δ, trattiamo ora di fasci di traiettorie rappresentati da funzioni $\tilde{\delta}$.

Gli insiemi di Gibbs possono essere considerati come superposizioni

Figura 5.8

Forma tipica di una densità δ_{ω_0} su una fibra che si contrae.

di punti dello spazio delle fasi, come

$$\rho(\omega) = \int_\Gamma \rho(\omega_0)\, \delta(\omega - \omega_0)\, d\omega_0 \tag{5.30}$$

Questa costruzione deve essere emendata nel caso dei sistemi dinamici instabili che vanno da ρ a $\tilde{\rho}$. Applicando Λ su entrambe le parti dell'equazione 5.30 vediamo che gli insiemi $\tilde{\rho}$ sono superposizioni delle funzioni $\tilde{\delta}$

$$\begin{aligned} \tilde{\rho}(\omega) &= \Lambda\rho(\omega) \\ &= \int_\Gamma \rho(\omega_0)\, \tilde{\delta}(\omega - \omega_0)\, d\omega_0 \end{aligned} \tag{5.31}$$

Questi insiemi possono essere chiamati *insiemi di Boltzmann*, poiché evolvono in accordo con i semigruppi di Markov, e portano a un aumento monotòno dell'entropia. La transizione dagli insiemi di Boltzmann a quelli di Gibbs corrisponde alla transizione tra dalla finitezza arbitraria alla precisione infinita. *Gli insiemi di Gibbs costituiscono pertanto un limite singolare, non fisico, dell'insieme di Boltzmann*. Sottolineiamo ancora una volta quanto questo differisca dall'atteggiamento che attribuisce l'irreversibilità a una approssimazione introdotta nella descrizione di processi irreversibili.

9. *La teoria cinetica.*

Abbiamo fin qui dedicato questo capitolo a una classe specifica di sistemi dinamici instabili, i flussi K e, particolarmente a un semplice esempio, la trasformazione del fornaio. Un certo numero di situazioni reali, come il moto browniano o il modello di Lorentz (cfr. sopra, § 1) possono essere studiate usando il metodo sviluppato qui.

Dedichiamo questo paragrafo a un'altra questione: perché l'irreversibilità è cosí comune in sistemi estesi, come i gas o i liquidi? Qual è l'origine dell'instabilità in tali sistemi? Questo interrogativo non è affatto banale. I modelli a sfere rigide, come il modello di Lorentz, non sono sistemi hamiltoniani, poiché il potenziale di energia di due sfere rigide è o infinito o zero, e non una funzione continua della distanza. Perciò l'equazione hamiltoniana del moto 2.5*b* non si applica. All'opposto in un sistema reale a molti corpi abbiamo interazioni senza salti e perciò una hamiltoniana contenente sia energia potenziale sia cinetica. Possiamo

cominciare a comprendere, come è nostro desiderio mostrare, perché questi sistemi estesi possono essere molto simili ai flussi K.

Torniamo al capitolo III; nel § 8 abbiamo rimarcato il ruolo delle condizioni di risonanza: nei sistemi a basso numero di dimensioni ivi considerati, per la maggior parte dei tori le frequenze sono incommensurabili (tori non risonanti); questo porta a un moto quasi-periodico attorno a punti ellittici. Abbiamo tuttavia anche una distribuzione densa di punti iperbolici che sorgono dalla distruzione dei tori risonanti. Vogliamo mostrare ora che, nei limiti dei sistemi estesi, è probabile che la situazione cambi drasticamente e che le condizioni di risonanza possano essere soddisfatte quasi ovunque.

Cominciamo osservando che in tali sistemi la dinamica può essere vista come un moto libero interrotto da incontri successivi tra particelle. Questa interazione è visibile molto chiaramente in un intorno del caso limite della particella libera, da una soluzione perturbativa formale dell'equazione di Liouville, in cui sono mantenuti solo i termini dominanti. Introducendo anche l'assunzione che le correlazioni iniziali tra particelle siano di corto raggio, arriviamo a un'equazione detta *equazione cinetica* che governa il rilassamento del sistema all'equilibrio. Questa equazione comprende un oggetto che gioca un ruolo centrale nella teoria cinetica moderna, l'*operatore di collisione*. Nell'ordine dominante della teoria della perturbazione questo operatore contiene un contributo moltiplicatore della forma

$$\left(\frac{2\pi}{L}\right)^3 \sum_{l} |V_l|^2 \boldsymbol{l}^2 \, \delta[\boldsymbol{l} \cdot (\boldsymbol{p}_j - \boldsymbol{p}_k)] \tag{5.32}$$

in cui $\boldsymbol{p}_j$, $\boldsymbol{p}_k$ sono i momenti delle particelle j e k, e δ è il delta di Dirac della funzione. V_l è la trasformata di Fourier dell'interazione potenziale V delle particelle j e k.

$$\begin{aligned} V(|\boldsymbol{x}_j - \boldsymbol{x}_k|) &= \left(\frac{2\pi}{L}\right)^3 \sum_{l} V_l e^{i\boldsymbol{l} \cdot (\boldsymbol{x}_j - \boldsymbol{x}_k)} \\ \boldsymbol{l} &= \frac{2\pi}{L} \boldsymbol{n} \end{aligned} \tag{5.33}$$

$\boldsymbol{n}$ essendo un vettore con componenti interi. Scrivendo quest'ultima relazione assumiamo condizioni al contorno periodiche, e che V dipenda solo dalla distanza $|\boldsymbol{x}_j - \boldsymbol{x}_k|$ tra le due particelle. Il fattore $(2\pi/L)^3$, in cui L^3 esprime il volume del sistema, assicura la transizione dalla serie di Fourier all'integrale di Fourier nei limiti di un sistema esteso, $L \to \infty$.

La validità della teoria cinetica implica che l'espressione esposta nella 5.32 esista e sia finita in un'intera regione dello spazio delle fasi. Analizziamo le conseguenze di questa proprietà per la dinamica. A causa della presenza del delta della finzione otteniamo contributi non evanescenti nell'equazione 5.32 per quegli $\boldsymbol{l}$ (con $|\boldsymbol{l}| \neq 0$) per cui

$$\boldsymbol{l} \cdot (\boldsymbol{p}_j - \boldsymbol{p}_k) = 0 \tag{5.34}$$

Qual è il significato di questa relazione? Dalla seconda delle equazioni 5.33, $\boldsymbol{l}$ è uguale (fino a un fattore moltiplicatore) a un vettore con componenti interi. In piú, in accordo con l'approccio perturbativo $\boldsymbol{p}_j$ e $\boldsymbol{p}_k$ sono associati con il moto delle particelle libere e perciò correlati all'inverso del tempo necessario alle particelle per attraversare la dimensione caratteristica L del sistema. In altre parole possono essere visti come analoghi alle frequenze nella rappresentazione azione-angolo. Ne segue che la 5.34 ha una struttura simile alla condizione di risonanza introdotta nel capitolo III. Come abbiamo visto in quel capitolo, § 8, le risonanze sono «pericolose», nel senso che ci si aspetta che una piccola perturbazione attorno al sistema integrabile di riferimento (qui il moto delle particelle libere) dia luogo al sorgere di punti iperbolici e moti instabili.

Finché $\boldsymbol{l}$ assume soltanto valori discreti l'equazione 5.34 sarà soddisfatta solo per certi $\boldsymbol{p}$. L'operatore di collisione cosí svanirà quasi ovunque nello spazio delle fasi. Tuttavia, nei limiti di un sistema esteso, $\boldsymbol{l}$ diviene una variabile continua (le equazioni 5.32 e 5.33, divengono integrali di Fourier). Come risultato ci aspetteremo che la risonanza appaia *quasi dappertutto* e che l'operatore di collisione esista e sia finito. Siamo perciò tentati di suggerire che nei sistemi in cui l'operatore di collisione non scompare, quasi tutti i punti dello spazio delle fasi sono iperbolici, come nella trasformazione del fornaio. Una volta ancora vediamo come l'irreversibilità, che è inerente alla teoria cinetica, può essere rinvenuta nell'instabilità e nella dinamica sottostante.

10. *Risonanza e interazione luce-materia.*

Abbiamo visto nel § 9 come le risonanze giochino un ruolo fondamentale nella teoria cinetica. Soltanto se possono essere soddisfatte quasi ovunque nello spazio delle fasi si può definire l'operatore di collisione, l'oggetto alla base della teoria moderna dei processi di non-

equilibrio. Le risonanze sono cosí responsabili dell'avvicinarsi all'equilibrio e dell'irreversibilità nel comportamento dei sistemi a molti corpi.

In questo paragrafo vedremo come le risonanze giochino un ruolo altrettanto fondamentale nell'interazione tra radiazione elettromagnetica e materia, una delle interazioni fondamentali in natura. Al contrario che nel § 9, tuttavia, dove le risonanze appaiono automaticamente non appena il sistema inizia a evolvere, scopriremo qui che esse si creano gradualmente come risultato di processi di riorganizzazione di nuovo tipo.

Consideriamo un oscillatore armonico di frequenza ω. La sua hamiltoniana è rappresentata nella equazione 2.6*b*. Supponiamo che l'oscillatore sia carico e incorporato in un campo magnetico con cui interagisce. Come è consueto nella teoria elettromagnetica, descriviamo il campo come un numero infinito di oscillatori armonici e assumiamo che la frequenza di questi sia strettamente distribuita intorno a qualche valore ω_e.

A causa della presenza del campo l'oscillatore subirà una forza proporzionale alla sua carica, dipendente dalla sua velocità e dall'intensità e orientazione del campo elettrico e magnetico. Questa forza determinerà un movimento accelerato, anche se inizialmente l'oscillatore è in quiete. Ora, nell'elettromagnetismo classico si mostra come una particella carica che subisce un moto accelerato, irradia, e pertanto perde, energia. La teoria che stabilisce questo «smorzamento di radiazione» non è completamente soddisfacente, e per un approccio piú fondamentale al fenomeno dell'emissione di radiazione dalla materia bisogna ricorrere alla meccanica quantistica. La seguente equazione di moto dà nondimeno, in prima approssimazione, un'idea del ruolo della risonanza nell'interazione luce-materia.

$$\frac{d^2x}{dt^2} + \frac{\gamma}{m}\frac{dx}{dt} + \omega^2 x = qF_e \cos \omega_e t \tag{5.35}$$

in cui x è la posizione coordinata dell'oscillatore, m e q sono rispettivamente la massa e la carica, γ è la costante di smorzamento e F_e è l'intensità del campo.

La soluzione della 5.35 può essere ottenuta con metodi standard nella forma

$$x = \rho F_e \cos(\omega_e t + \theta) \tag{5.36}$$

con

$$\rho^2 = \frac{q^2}{(\omega^2 - \omega_e^2)^2 + \frac{\gamma^2}{m^2}\omega_e^2} \tag{5.37a}$$

$$\tan \theta = \frac{\gamma\omega_e}{m(\omega^2 - \omega_e^2)} \tag{5.37b}$$

Osserviamo un'estensione molto forte della risposta quando la frequenza intrinseca ω dell'oscillatore si avvicina a quella del campo ω_e. Pertanto, nei limiti di un modesto smorzamento γ, l'oscillatore irradierà essenzialmente alla frequenza del campo. Questo è precisamente il fenomeno della risonanza.

Guardiamo ora il problema da un punto di vista piú fondamentale. Consideriamo l'intera hamiltoniana dell'oscillatore carico piú gli oscillatori di campo, inclusa la loro interazione. Iniziamo con una condizione iniziale corrispondente all'oscillatore carico in moto e gli oscillatori di campo in quiete, o equivalentemente, allo stato di energia zero (questo è il vuoto classico, che, emendato dalla meccanica quantistica, corrisponde effettivamente allo stato non zero di energia). Risolviamo poi al calcolatore l'equazione totale (reversibile) di moto, senza introdurre l'assunzione fenomenologica dell'esistenza di una costante di smorzamento γ. Le simulazioni mostrano che vi è un breve intervallo prima che l'oscillatore carico inizi a irradiare, cosa di cui l'equazione 5.35 non può dare adeguatamente conto. È molto interessante comprendere l'origine di questo ritardo. Abbiamo appena visto come il processo di radiazione implichi una risonanza tra l'oscillatore materiale e gli oscillatori di campo; ma all'inizio non vi sono oscillatori di campo disponibili: come primo passo l'oscillatore materiale deve pertanto produrre gli oscillatori di campo, o, in altre parole, creare l'ambiente giusto per essere poi in grado di irradiare. Quali che siano le condizioni iniziali, dopo un breve intervallo (nell'ordine compreso tra 10^{-14} e 10^{-18} secondi per l'atomo d'idrogeno), l'oscillatore inizia a irradiare. In breve il meccanismo della radiazione è un processo irreversibile «autocatalitico» che compare nell'interfaccia tra il vuoto e la materia.

Paragoniamo ora questo comportamento con i processi irreversibili nei fluidi. Anche qui, quale che sia la preparazione, è impossibile impedire al fluido di raggiungere l'equilibrio termico. In quest'ultimo caso, tuttavia, la scala temporale è dell'ordine del tempo di rilassamento (normalmente 10^{-9} secondi), che è molto piú lungo del tempo caratte-

ristico dei processi irreversibili che preparano l'emissione della radiazione.

L'esistenza di processi di rilassamento in scale temporali brevissime, associata con l'interazione della radiazione con la materia illustra ancora una volta il ruolo universale dell'irreversibilità in natura.

11. *Osservazioni conclusive.*

Viviamo in un mondo pluralistico. Fenomeni dinamici semplici, come il moto periodico della Terra attorno al Sole ci appaiono come temporalmente reversibili e deterministici. Ma troviamo anche processi irreversibili, come le reazioni chimiche, e processi stocastici come la scelta tra stati di biforcazione.

Com'è possibile tutto ciò? Secondo la concezione classica l'irreversibilità e la casualità su scala macroscopica erano apparenze dovute alla complessità del comportamento collettivo di oggetti intrinsecamente semplici. La credenza nella semplicità del livello fondamentale è stata una delle forze dominanti della scienza classica degli ultimi tre secoli.

Oggi dobbiamo riconsiderare la situazione. Il livello fondamentale non è semplice, neppure nella fisica classica, che per tanto tempo è stata il bastione del determinismo e della reversibilità. Il punto di vista classico emerse dallo studio di sistemi in cui il moto periodico e i punti ellittici giocavano un ruolo essenziale. All'opposto, vediamo ora che esiste una grande quantità di sistemi in cui sono centrali i punti iperbolici. In questo capitolo abbiamo indicato due tipi di sistemi, i flussi K e i sistemi studiati nella teoria cinetica moderna, in cui quasi tutti i punti nello spazio delle fasi sono iperbolici. Per questi sistemi il concetto di traiettoria perde ogni sostegno osservativo, e la loro dinamica deve essere incorporata in un formalismo piú ampio che comprenda la stocasticità e l'irreversibilità. Questo passaggio, che porta dai gruppi ai semigruppi, non è dovuto alla nostra «ignoranza», ma alla stessa struttura dei sistemi dinamici.

In fisica è spesso utile distinguere tra stati e leggi. Gli stati possono essere conosciuti solo con la preparazione e l'osservazione di un sistema, e la specificazione di questi processi è una parte indispensabile di ogni descrizione oggettiva del sistema, come rimarcavano, in particolare, Niels Bohr e Leon Rosenfeld. Questa preparazione e osservazione può portare solo a una «finestra finita». Il meccanismo delle nostre misurazioni non può differire da ogni altra interazione naturale tra sistemi

dinamici, dunque comprende la finestra finita associata con la loro limitata sensitività.

Il riconoscimento di questa finestra finita porta a un nuovo rapporto tra stati e leggi. Nella concezione classica (inclusa la meccanica quantistica) gli stati sono temporalmente simmetrici, e sono propagati da leggi che conservano la simmetria temporale. Ora dobbiamo prendere in considerazione stati come quelli degli insiemi di Boltzmann, esposti nel § 8, che hanno rottura di simmetria temporale e che sono propagati da leggi dotate di rottura di simmetria temporale.

Capitolo sesto

La complessità e il trasferimento di conoscenza

Comprendere cosa avviene attorno a noi equivale a costruire dei modelli e a confrontarli con le osservazioni. Questa affermazione può suonare come una verità evidente a un fisico o a un chimico, ma va ben oltre la fisica e la chimica. In ogni momento il nostro cervello esamina l'ambiente circostante, classifica le osservazioni rispetto alle immagini già formate, e alla fine, raggiunge una conclusione preliminare. Uno dei fondamentali passi in questa procedura è l'ampio uso di *archetipi* e di *analogie*.

Il nostro scopo in questo capitolo è di mostrare che i sistemi fisico-chimici che dànno origine ai fenomeni di transizione, all'ordine a lungo raggio, e alla rottura della simmetria lontano dall'equilibrio possono servire come un archetipo per capire altri tipi di sistemi che mostrano un comportamento complesso, per i quali le leggi di evoluzione delle variabili coinvolte non sono note ad alcun livello confrontabile di dettaglio. Ancora piú importante, in molti di questi sistemi la scelta di quella che deve essere una variabile pertinente può ben essere parte del problema che uno cerca di risolvere.

In tutti gli esempi discussi, l'analisi procede in due passi. Primo, vengono tratte certe analogie fra le osservazioni e il comportamento di sistemi fisico-chimici di «riferimento». Ciò definisce il tipo di modello che è probabilmente la piú adeguata rappresentazione del sistema di interesse. Poi si prova ad andare oltre lo stadio delle semplici analogie, per riconoscere, *entro il quadro del modello adottato*, le specificità di ogni problema, e di incorporarle nella descrizione. Le previsioni dell'analisi vengono infine confrontate con l'esperienza relativa al comportamento nel passato e, se si raggiunge un accordo qualitativo, esse vengono utilizzate per prevedere le tendenze future. Se tutto procede bene, come naturale risultato di questo processo dovrebbero emergere suggerimenti concreti su come *dominare la complessità*.

1. *La dinamica non-lineare in condizioni lontane dall'equilibrio e la modellistica della complessità.*

Nel corso di questo libro la complessità è stata connessa alla capacità di cambiare fra diversi modi di comportamento al variare delle condizioni dell'ambiente circostante. La flessibilità e l'adattabilità dei risultati introducono, a loro volta, la nozione di *scelta* fra le varie possibilità offerte. È stato sottolineato che la scelta è mediata dalla dinamica delle fluttuazioni e che richiede l'intervento delle loro due manifestazioni antagoniste: casualità a corto raggio, che fornisce l'elemento «innovativo» necessario per esplorare lo spazio degli stati; e l'ordine a lungo raggio, che mette in grado il sistema di mantenere un regime collettivo che comprende le regioni spaziali macroscopiche e gli intervalli di tempo macroscopici.

Un prerequisito necessario di tutti questi fenomeni è la dinamica non-lineare, che dà luogo, sotto opportuni vincoli, all'*instabilità del moto* e alle *biforcazioni*. Il primo passo per creare i modelli per il comportamento complesso, è dunque l'accertamento del carattere non-lineare della dinamica sottostante e l'identificazione di un insieme di variabili in grado di mostrare le instabilità e le biforcazioni. Questa procedura è molto facilitata dai risultati della teoria dei sistemi dinamici esaminata nel capitolo III (specialmente il § 6), secondo cui l'essenza del comportamento qualitativo è spesso contenuta in semplici equazioni di evoluzione universali, le *forme normali*, soddisfatte da un insieme limitato di variabili collettive, i *parametri d'ordine*. Potenzialmente, dunque, un sistema a molte variabili, come parecchi fra i sistemi incontrati in pratica, si possono ridurre a una descrizione semplificata vicino al punto di biforcazione. Per esempio, se per qualche ragione sappiamo che il sistema in considerazione può subire transizioni verso due stati stazionari simultaneamente stabili senza una rottura della simmetria spaziale, siamo tentati di usare una dinamica cubica (cfr. equazione 3.14), per una descrizione di ordine zero. Se, d'altra parte, siamo in presenza di una transizione verso un comportamento oscillatorio forzato, saremmo autorizzati a guardare il sistema delle equazioni 3.16 per le variabili radicali e di fase come un promettente punto di partenza. Infine, una dinamica caotica suggerirebbe che una descrizione continua debba coinvolgere almeno tre variabili e che, in una descrizione discreta data, per esempio, dalla mappa di Poincaré, si possa cercare una dinamica iterativa simile a quella rappresentata dalla figura 3.27 e dalle equazioni 3.31.

Queste considerazioni, la cui utilità non deve essere sottovalutata, devono naturalmente essere applicate con lucidità, altrimenti possono oscurare ciò che è, e deve rimanere, un aspetto importante del problema, ossia la specificità del sistema considerato. A ogni stadio dell'analisi il «modellatore» deve dunque essere capace di collegare i parametri d'ordine alle variabili originali presenti, di annettere la forma dinamica normale al meccanismo dei processi che si svolgono nel sistema, e di soddisfare i vincoli che possono essere imposti dalla natura propria del problema. Nei sistemi riconducibili alle variabili fisico-chimiche, questi possono essere la positività della produzione di entropia, della temperatura o delle concentrazioni. Nei problemi delle popolazioni, analogamente, i vincoli imporranno che la popolazione di una specie non possa mai essere negativa. E nei problemi geofisici essi selezioneranno modelli in cui le scale spaziali e temporali riportate dai dati siano adeguatamente riprodotte dalla forma delle funzioni di velocità dai valori dei parametri adottati.

Tenendo presenti questi concetti procediamo ora a illustrare una serie di esempi in cui viene applicata con successo la modellistica che usa come archetipo i sistemi dinamici non-lineari lontano dall'equilibrio. Consideriamo, successivamente, la scienza dei materiali, la biologia, la geofisica e il comportamento sociale delle popolazioni umane e animali.

2. *La scienza dei materiali.*

Un campo in cui l'impatto degli sviluppi della dinamica non-lineare in condizioni lontane dall'equilibrio sta rapidamente crescendo, è la scienza dei materiali. La sintesi di nuovi materiali è una parte cruciale dei nuovi sviluppi tecnologici. È sorprendente scoprire come, in molti casi, essa sia condizionata da processi che operano a una grande distanza dall'equilibrio termodinamico.

Per esempio si è oggi compreso che i fenomeni di trasporto che avvengono nella fase fluida ospite, durante il processo di crescita dei cristalli e la solidificazione delle leghe, hanno una profonda influenza sulla struttura e sulle qualità della fase solida. Come in molti problemi che coinvolgono i fluidi sotto i vincoli di non-equilibrio, la convezione gioca un ruolo dominante perché influisce sulla temperatura e sulla composizione all'interfaccia fra le fasi solida e liquida. Il movimento convettivo del fluido è a sua volta determinato dalla competizione fra condizioni esterne (convezione forzata) e processi interni (convezione naturale).

Questi ultimi sono dovuti a differenze di concentrazione o di temperatura (cfr. capitolo I, § 3), ma possono anche essere indotte dalla tensione superficiale (cfr. capitolo I, § 5). La crescita dei cristalli e la solidificazione sono processi accoppiati che implicano trasferimento di massa e di calore, flussi di fluido, reazioni chimiche, e transizioni di fase. Come sottolineato nel capitolo I, § 3, e nel corso del capitolo III, la convezione può diventare irregolare sotto certe condizioni. Questo comportamento caotico induce variazioni casuali di temperatura e di concentrazione, che risultano in imperfezioni della struttura cristallina.

Le proprietà dell'interfaccia fra solido e liquido sono naturalmente di primaria importanza in questi fenomeni. Una manifestazione sorprendente del non-equilibrio in questo contesto è la scoperta che durante la crescita del cristallo esiste una fase omogenea fra solido e liquido, le cui proprietà sono diverse da quelle della sostanza fusa e del cristallo. Gli studi eseguiti con la diffusione della luce su questo strato, le cui dimensioni sono sostanzialmente piú grandi del passo reticolare, rivelano un coefficiente di diffusività termica dell'ordine di 10^{-8} cm^2, in contrapposizione al valore caratteristico di 10^{-3} cm^2 sec^{-1} per l'acqua pura. In altre parole, i disturbi termici vengono smorzati in modo molto meno efficiente, proprietà questa che suggerisce come l'intero strato sia in condizioni critiche simili a quelle prevalenti vicino a un punto di biforcazione. Presumibilmente questa transizione è mediata dal flusso di calore attraverso l'interfaccia che nasce dall'avanzamento del fronte di solidificazione, che letteralmente mantiene nello strato una *fase di nonequilibrio della materia* (detta anche «mesofase»).

Un'altra classe di fenomeni di non-equilibrio nasce con l'irraggiamento da potenti laser o sorgenti a particelle, che permettono di iniettare grandi quantità di energia nello strato superficiale di vari materiali in un intervallo di tempo molto breve. I fenomeni di fusione e di risolidificazione che ne risultano sono caratterizzati da alti valori di gradienti termici (da 10^6 a 10^8 K cm^{-1}) e dalle velocità di interfaccia (1 m sec^{-1}). Queste condizioni estreme di non-equilibrio promettono, fra l'altro, la sintesi di nuovi materiali elettronici.

Forse il piú sorprendente intervento del non-equilibrio nel campo della scienza dei materiali riguarda la degradazione dei materiali. Fenomeni come la plasticità, la deformazione o la fatica sono generalmente veicolati da difetti generati dalle tensioni, che successivamente si amplificano e si diffondono per formare strutture che precedono la rottura. Il progredire in questo campo richiede una comprensione della formazione dei difetti, delle dislocazioni, e delle loro interazioni nei siste-

mi soggetti a vincoli esterni. Una modellistica delle dinamiche non-lineari risultanti condurrebbe a instabilità che inducono transizioni multiple a stato stazionario e strutture spaziali per la concentrazione del difetto. Descriviamo ora a grandi linee alcuni dei punti principali.

Identifichiamo dapprima le variabili e i parametri pertinenti. Oltre alla deformazione ε, che è la variabile generalmente presente nell'approccio ingegneristico, anche la temperatura T sembra giocare un ruolo. Infatti, durante una prova di tensione un campione si raffredda mentre è sottoposto alla trasformazione nella regione elastica. La quantità di temperatura ceduta è proporzionale alla deformazione (che per semplicità assumiamo avvenire omogeneamente all'interno della massa) e dipende dal *parametro di Grüneisen*, γ. Quest'ultima quantità rispecchia la deviazione delle forze intermolecolari dal comportamento dell'oscillatore armonico e fornisce quindi una misura dell'accoppiamento termomeccanico. A temperature ordinarie γ varia tra 1,5 e 2,5 per una ampia classe di materiali, ma sembra anche dipendere fortemente dalla tensione, approssimandosi a un valore praticamente nullo per una grandezza critica del vincolo applicato.

D'altra parte alla diminuzione di temperatura ora citata si oppongono gli effetti di riscaldamento collegati alla viscosità del materiale. Non appena la tensione applicata aumenta, le dislocazioni inizialmente statiche e gli altri difetti divengono sempre piú mobili, e il loro moto all'interno del campione produce una quantità crescente di energia termica. Questo, a sua volta, accresce ulteriormente la popolazione dei difetti, non diversamente da ciò che accade durante una reazione chimica esotermica (cfr. capitolo 1, § 4). Siamo quindi in presenza di un meccanismo destabilizzante con *feedback* positivo, che controbilancia l'effetto stabilizzante del raffreddamento indotto dalla deformazione. I due meccanismi si bilanciano l'un l'altro in un qualche punto critico, che può essere interpretato come una frontiera tra i regimi elastico e plastico della deformazione. La figura 6.1 rappresenta questa notevole coincidenza. Ci si aspetta che il valore di soglia, σ_θ, della tensione – poiché il vincolo esterno conduce il sistema fuori dall'equilibrio – in questa situazione critica, definisca un punto di biforcazione di un nuovo stato.

Una descrizione piú quantitativa del fenomeno precedente richiede la deduzione di un insieme di equazioni di evoluzione per le variabili chiave ε e T, nelle quali l'effetto dei parametri σ e γ può essere trovata esplicitamente. Non andiamo nei dettagli, ma notiamo solo che sotto ipotesi ragionevoli queste due equazioni risultano essere accoppiate non-linearmente, attraverso il termine

$$-\frac{1}{2}\,C_v\left(\frac{\partial\gamma}{\partial\varepsilon}\right)\left(\frac{T-T_0}{T_0}\right)\varepsilon^2 \qquad (6.1)$$

dove C_v è il calore specifico e T_0 una temperatura di riferimento. A causa della dipendenza dalla tensione del parametro di Grüneisen γ, notata sopra, questo termine è positivo e introduce un meccanismo di accrescimento delle fluttuazioni che conduce alla instabilità. Vicino al punto di biforcazione σ_θ, seguendo i risultati del capitolo III, § 6, si può semplificare ulteriormente la descrizione eliminando tutte le variabili eccetto il parametro d'ordine. Quest'ultimo, che è una opportuna combinazione lineare di ε e $T-T_0$, segue una dinamica di forma simile all'equazione 3.14 o alla prima delle equazioni 3.16:

$$\frac{dx}{dt} = (\sigma-\sigma_\theta)x - bx^3 \qquad (6.2)$$

A seconda delle condizioni, la biforcazione può condurre sia a un nuovo stato stazionario (come nell'equazione 3.14) o a una oscillazione di ciclo limite (come nell'equazione 3.16).

Un approccio piú microscopico al precedente problema richiede l'analisi del bilancio tra differenti processi elementari come la creazione, l'annientamento, l'ammassamento e la diffusione delle dislocazioni.

Figura 6.1

Variabili e fenomeni in funzione del tempo per un campione di acciaio: *a*) tensione; *b*) deformazione; *c*) temperatura; *d*) attività di emissione acustica. Si noti il brusco aumento della temperatura e dell'attività acustica nella regione critica.

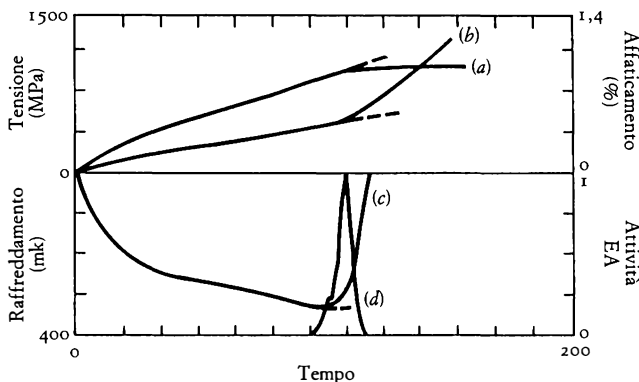
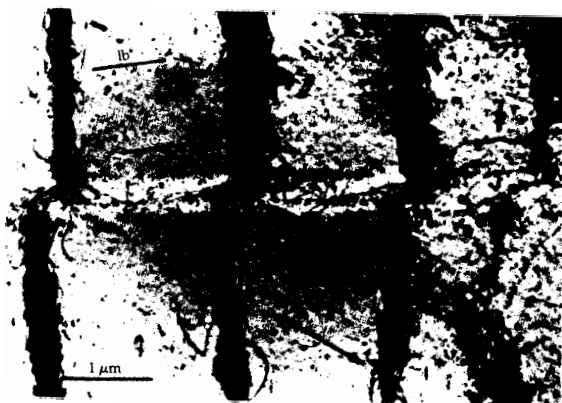
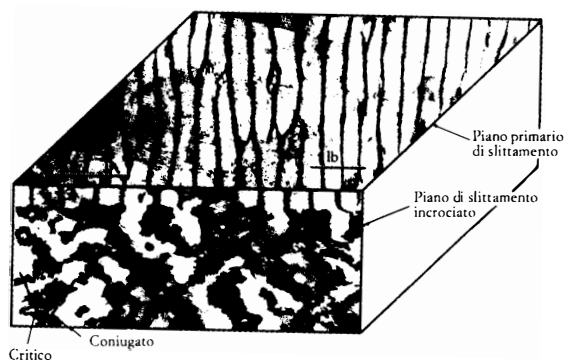
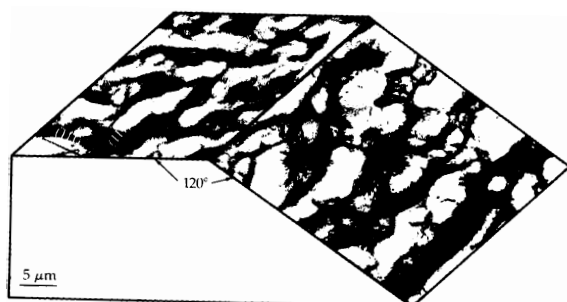



Figura 6.2

La struttura di dislocazione in tre dimensioni in un singolo cristallo di rame periodicamente sotto tensione.

Questo porta a modelli di reazione-diffusione molto simili a quelli usati per studiare le instabilità chimiche (cfr. capitolo III, § 11). Quando questi modelli sono soggetti a vincoli anisotropi, prevedono una gerarchia di transizioni con rottura di simmetria che conduce a strutture di dislocazione molto simili a quelle osservate sperimentalmente (cfr. figura 6.2).

Concludendo, sembra che i fenomeni meccanici di tale attuale importanza e interesse come la plasticità e la flessibilità non possano essere investigati nell'unico contesto della meccanica. Piuttosto, debbono essere considerati come parti della problematica generale dei sistemi dinamici non-lineari che agiscono in condizioni lontane dall'equilibrio. È nostra convinzione che questa idea costituisca per se stessa un importante passo avanti nel campo della scienza dei materiali.

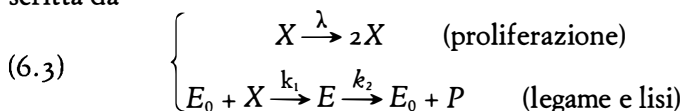
3. *Fenomeni di soglia nella dinamica cellulare.*

Passare dai materiali sotto sforzo alle cellule viventi può sembrare un'impresa aleatoria, ma il nostro lettore dovrebbe avere ormai compreso che tale processo è al centro dell'arte della modellistica. Abbiamo appena visto che la plasticità e la deformabilità si possono considerare come un fenomeno di soglia mediato dalle interazioni competitive fra la popolazione dei difetti e la massa materiale. Quando si verifica la deformazione, dobbiamo dunque concludere che il materiale non è stato in grado di controllare l'attività dei difetti, la cui popolazione ha raggiunto come risultato livelli catastrofici. Scopo del presente capitolo è mostrare che una competizione molto simile si verifica anche in fenomeni biologici di vitale importanza come, per esempio, la crescita di tumori. Qui una piccola popolazione iniziale di «provocatori», per esempio certe cellule di un tessuto, ha perso, per alcune ragioni rivelate dalla biologia molecolare, la sua funzione fisiologica e diventa maligna. Essa tende successivamente a invadere l'organismo per rapida proliferazione, ma l'organismo cerca di controbattere la sua azione mandando sul «campo di battaglia» cellule «killer» specializzate. Il risultato della conseguente competizione tra cellule maligne e cellule «killer» deciderà se i provocatori saranno respinti o se, al contrario, domineranno. L'analisi di questa competizione rivela quello che crediamo essere un aspetto della piú grande importanza, precisamente che il rigetto o il dominio può essere visto come un fenomeno di soglia.

Prima di delineare la competizione in termini piú quantitativi, diamo una breve descrizione della natura delle cellule «killer». Il sistema

immunitario, che fornisce i mezzi con i quali i vertebrati possono controbattere l'azione di sostanze patogene che invadono l'organismo, produce le cellule non differenziate del ceppo immune nel midollo osseo. Queste cellule si differenziano successivamente in linfociti di tipo B o T. Nello scontro con una determinata sostanza invadente, di solito chiamata l'*antigene*, le cellule B si differenziano ulteriormente in cellule piú grandi che proliferano e secernono sostanze chimiche capaci di neutralizzare l'antigene, conosciute come *anticorpi*. Le cellule T, d'altro canto, dopo ulteriori differenziazioni che avvengono nel timo, regolano, sia con l'accrescimento che con la soppressione, l'azione delle cellule B. Inoltre esse sono coinvolte nelle risposte immunitarie che sono direttamente mediate dalle cellule, particolarmente contro le cellule cancerogene e le cellule estranee che provengono, per esempio, dai trapianti. Altre specie cellulari, come i macrofagi, hanno una simile attività citotossica. È questa risposta mediata dalle cellule che ci interessa principalmente in questo paragrafo.

Sia X la densità di popolazione delle cellule bersaglio (maligne) che prolificano. E_0 quella delle cellule citotossiche libere. In accordo con i dati sperimentali, le cellule E_0 riconoscono le X, le fissano nella forma di un complesso $E \equiv E_0X$ e successivamente le lisano attraverso la dissociazione del complesso E. Questa successione di passi può essere descritta da

$$(6.3) \qquad \begin{cases} X \xrightarrow{\lambda} 2X & \text{(proliferazione)} \\ E_0 + X \xrightarrow{k_1} E \xrightarrow{k_2} E_0 + P & \text{(legame e lisi)} \end{cases}$$

Assumiamo che nell'intervallo di tempo della competizione $E_0 + X$, lo stato del sistema immunitario si possa considerare come stazionario a tutti gli effetti pratici. Questo significa che la densità totale di popolazione $E_t = E_0 + E$ rimane costante nel tempo. Adottiamo inoltre una descrizione semplificata per la quale la cinetica dei due passi nella seconda equazione 6.3 si può considerare come non cooperativa (senza *feedback*), almeno nell'intervallo dei valori dei parametri che ci interessa. Enfatizziamo comunque che questa ipotesi si può facilmente indebolire senza influenzare in modo qualitativo i risultati.

In queste condizioni possiamo scrivere le seguenti equazioni del bilancio:

$$dX/dt = \lambda X[1 - (X/N)] - k_1 E_0 X$$

$$dE_0/dt = -k_1 E_0 X + k_2 E$$

nella prima delle quali il fattore $(1 - X/N)$ esprime l'esistenza di un limite superiore N per X all'interno dell'elemento di volume dove si svolge la competizione. Poiché ci si attende che la lisi sia molto piú veloce di altri processi, l'approssimazione di stato quasi stazionario illustrata nel capitolo III, § 6, può essere invocata per la seconda equazione. Ponendo $dE_0/dt \sim 0$ e usando la condizione di conservazione delle specie E, siamo in grado di esprimere E_0 in termini di X e ottenere

$$\frac{dX}{dt} = \left(1 - \frac{X}{N}\right)\lambda x - \frac{k_1 E_t X}{1 + \frac{k_1}{k_2} X} \tag{6.4}$$

Questa equazione si può riscrivere in forma adimensionale come:

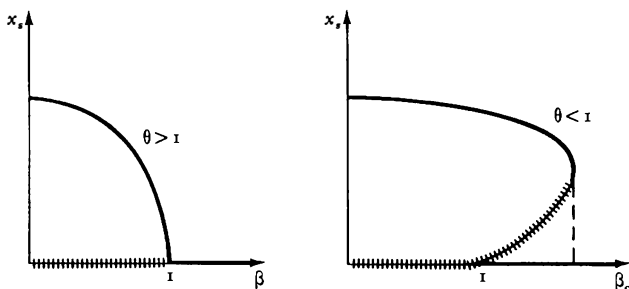
$$\frac{dx}{d\tau} = (1 - \theta x)x - \frac{\beta x}{1 + x} \tag{6.5}$$

dove:

$$\beta = \frac{k_1 E_t}{\lambda} \qquad \theta = \frac{k_2}{k_1 N} \qquad \tau = \lambda t \qquad x = \frac{k_1}{k_2} X \tag{6.6}$$

È ovvio che l'equazione 6.5 ammette sempre la soluzione di stato stazionario $x = 0$. L'analisi della stabilità lineare (cfr. *Appendice* A) mostra che per $\beta < 1$ questa soluzione è instabile, mentre per $\beta > 1$ è stabile, indipendentemente dal valore di θ. L'esistenza di una soluzione non nulla di stato stazionario per x corrisponde agli stati tumorali del tessu-

Figura 6.3

Concentrazione di uno stato stazionario di cellule tumorali in funzione del parametro β che caratterizza l'attività del sistema immunitario. Nella parte destra si riconosce la possibilità di coesistenza di uno stato normale ($x_S = 0$) e maligno ($x_S > 0$) separati da una soglia finita.

to. Due situazioni qualitativamente diverse possono quindi nascere, a seconda del valore di θ. Esse sono rappresentate nella figura 6.3. Per $\theta > 1$, l'esistenza di uno stato tumorale è ristretto ai valori di $\beta < 1$. La condizione per il rigetto è semplicemente che nel corso della reazione immunitaria il valore di β diventi piú grande di 1. Per $\theta < 1$, la transizione tra situazioni tumorali e non tumorali, coinvolge un fenomeno di bistabilità le cui proprietà ricordano le transizioni di non-equilibrio discusse nel corso di questo libro. Nel dominio $1 < \beta < \beta_c = (1 + \theta)^2/4\theta$, lo stato nullo (normale) e lo stato tumorale sono simultaneamente stabili. Ci si aspetta che la transizione fra questi stati coinvolga un processo di nucleazione. Cominciando dalla condizione iniziale di stato tumorale e incrementando il valore di β, il rigetto è assicurato solo per $\beta > \beta_c$.

Una stima dei valori del parametro ottenuta dalla letteratura ci porta alla conclusione che nella realtà entrambi i tipi di stato stazionario possano nascere. Ricordando dalla prima relazione 6.6 che il parametro β descrive essenzialmente l'efficienza del sistema immunitario, comprendiamo che per un dato stato del sistema immunitario nell'intervallo $1 < \beta < \beta_c, \theta < 1$, possono nascere entrambe le situazioni, normale e patologica, a seconda della storia del sistema.

Questa visione dinamica dovrebbe fornire un importante elemento della comprensione di una moltitudine di altri fenomeni biologici a livello sovramolecolare, come per esempio, l'aggregarsi delle cellule del sangue nella arteriosclerosi.

4. *Il modello del mutamento climatico e della variabilità.*

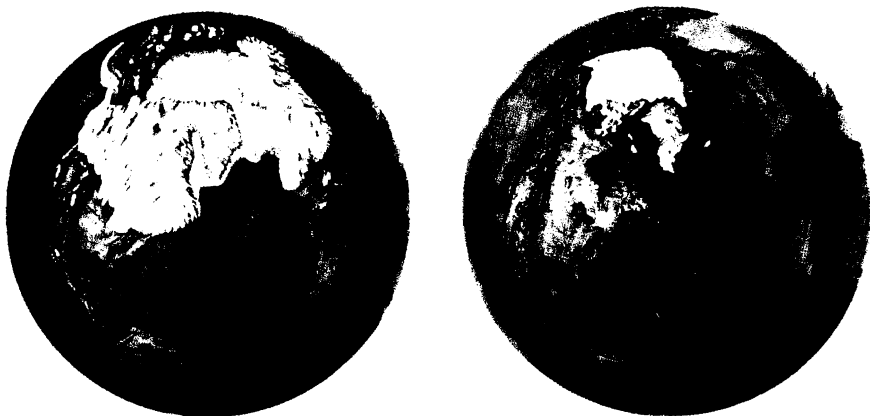
Le popolazioni di entità interagenti saranno al centro delle nostre discussioni piú avanti nel capitolo in rapporto con il modello del comportamento sociale. Prima di arrivare a questo problema, tuttavia, presentiamo qualche idea riguardante il modello della variabilità climatica, un problema di ovvia importanza non solo a livello scientifico, ma anche a livello economico e sociale. Tuttavia, nonostante questo carattere globale, la variabilità climatica rimane fondamentalmente un problema che si può ricondurre a descrizioni fisico-chimiche, in modo molto simile a quanto discusso piú sopra nel presente capitolo.

I fatti principali che suggeriscono come il sistema climatico sia caratterizzato da una variabilità pronunciata sono stati esaminati nel capitolo 1, § 8. La figura 6.4 mostra il quadro di ciò che sarà lo scopo principale del presente paragrafo, ovvero, la modellistica di questo fenomeno.

Nell'emisfero nord al massimo della glaciazione principale, 18 000 anni fa, uno strato di ghiaccio spesso parecchi chilometri copre buona parte dell'America del Nord, fino a metà degli Stati Uniti, come anche buona parte dell'Europa settentrionale e occidentale, fino alla latitudine di Parigi e Berlino. Nella situazione attuale il ghiaccio continentale è essenzialmente confinato alla Groenlandia. Inoltre, è noto che per quanto riguarda l'estensione del ghiaccio continentale, l'attuale situazione era praticamente stabilita già 10 000 anni fa. In altre parole siamo indotti a concludere che il pianeta Terra, un sistema fisico, ha subito, in poche migliaia di anni (un tempo breve nella scala geologica) una transizione fra due stati straordinariamente differenti, la cui diversità si estende alle dimensioni stesse della Terra.

Le osservazioni di cui sopra suggeriscono che ogni ragionevole modello del sistema climatico debba rendere conto della possibilità di tali transizioni su larga scala. Esaminiamo ora gli elementi che abbiamo a disposizione per costruire un modello adeguato di questo tipo.

I mutamenti climatici comprendono essenzialmente la superficie solida della Terra, gli oceani (l'idrosfera) e i ghiacciai (la criosfera). Lo stato di ciascuno di questi sottosistemi, come le loro interazioni, è mediato da fenomeni fisici che soddisfano le fondamentali equazioni del bilancio dell'energia, della massa e del momento. Il bilancio di energia de-

Figura 6.4
Illustrazione schematica dell'emisfero nord: al massimo dell'ultima glaciazione (a sinistra) e nelle attuali condizioni (a destra).

scrive come la radiazione solare in arrivo venga trasformata dal sistema: quali sono le parti che vengono assorbite e riemesse, come vengono distribuite alle differenti latitudini e longitudini cosí come sulle superfici continentali e oceaniche. Le parti piú significative del bilancio di massa sono, d'altra parte, il bilancio del ghiaccio continentale e marino, e di alcuni costituenti dell'atmosfera, minori ma fondamentali, come l'anidride carbonica e il vapore acqueo. Per ultimo, il bilancio del momento determina circolazioni atmosferiche, le quali alla fine condizionano la temperatura locale e il quadro delle precipitazioni. Il problema risiede nel fatto che questi bilanci sono intimamente accoppiati. Inoltre, la forma precisa delle leggi fenomenologiche che esprimono i vari flussi presenti in termini di un insieme limitato di variabili macroscopiche (cfr. capitolo II, § 2) è scarsamente conosciuto. Inoltre una delle poche realtà indiscusse è l'esistenza di turbolenze atmosferiche e oceaniche. Se non altro, ciò suggerisce (cfr. capitolo III, § 10) che il clima dovrebbe condividere l'impredicibilità intrinseca dei sistemi soggetti a dinamica caotica.

Di fronte a questa situazione si possono assumere due atteggiamenti. Da un lato si può sostenere che la totalità delle equazioni del bilancio, per le quali ogni flusso individuale è parametrizzato nel miglior modo possibile a seconda dei dati disponibili, debba essere risolta numericamente con un calcolatore. Questa prospettiva diede origine ai *modelli generali di circolazione* e stimolò lo sviluppo di calcolatori ultraveloci in grado di maneggiare questa enorme quantità di dati immessi.

Dal lato opposto si può argomentare che, proprio a causa dell'enorme complessità del sistema climatico, è necessaria una analisi *qualitativa* che abbia lo scopo di estrapolare tendenze a lungo termine e di identificare i parametri chiave. Per il lettore di questo libro questo atteggiamento dovrebbe apparire infatti come il naturale strumento per trattare la complessità. Dopo tutto, la maggior parte dei progressi nella comprensione dei fenomeni a livello di laboratorio, come le instabilità idrodinamiche e chimiche, che sono sostanzialmente piú semplici e piú facili da controllare del clima globale, è dovuta all'uso estensivo dei concetti e dei metodi della teoria dei sistemi dinamici. Nel resto di questo paragrafo daremo pertanto alcuni esempi sul tipo di informazione e di introspezione offerti dall'analisi qualitativa di semplici modelli matematici del clima. Focalizzeremo l'attenzione sul problema della glaciazione nell'èra Quaternaria e, in particolare sui meccanismi per i quali il sistema climatico può catturare e amplificare le piú deboli variazioni dell'energia solare in arrivo. Come abbiamo visto nel capitolo I, § 8, no-

nostante i loro piccoli valori, queste variazioni sembrano presentare una impressionante correlazione con le scale temporali caratteristiche delle glaciazioni.

Discutiamo ora in termini molto globali considerando la Terra come un punto nello spazio che riceve radiazione solare e che riemette nello spazio radiazione infrarossa. In quest'ottica vi è solo una variabile di stato importante, la temperatura media T, che evolve nel tempo secondo l'equazione del bilancio del calore

$$\frac{dT}{dt} = \frac{1}{C}\{Q[1 - a(T)] - \varepsilon_B \sigma T^4\} \tag{6.7}$$

C indica la capacità termica del sistema, Q la costante solare e a l'albedo, che esprime la parte delle radiazioni solari riemessa nello spazio. La dipendenza dalla temperatura di questa quantità tiene conto del *feedback* della albedo superficiale descritta alla fine del capitolo I, § 7, e può essere grossolanamente descritta dalla curva a della figura 6.5.

L'ultimo termine dell'equazione 6.7 descrive la radiazione termica emessa come una correzione alla ben nota legge di Stefan-Boltzmann della radiazione del corpo nero. σ è la costante di Stefan e ε_B è un fattore di emissività che tiene conto del fatto che la Terra non irraggia esattamente come un corpo nero.

Lo studio dell'equazione 6.7, riassunta nella figura 6.5, mostra che il sistema può ammettere fino a tre stati stazionari. Due di essi, denotati con T_b e T_a, sono stabili e corrispondono rispettivamente al clima attuale e ad un clima freddo che ricorda una glaciazione globale. Il terzo stato T_0 è instabile e separa i precedenti due regimi stabili.

Come abbiamo visto nel capitolo IV, § 5, l'evoluzione di un sistema che coinvolge solo una variabile si può sempre rappresentare come l'evoluzione di una «particella» in un potenziale U che, nel presente caso, è definito da (cfr. equazione 4.30)

$$U(T) = -\frac{1}{C}\int dT\{Q[1 - a(T)] - \varepsilon_B \sigma T^4\}$$

Chiamiamo $U(T)$ il *potenziale cinetico*. Sotto le condizioni della figura 6.5, questo potenziale è bistabile, proprio come in figura 4.7*b*.

Ora il clima, come ogni altro sistema fisico, è continuamente soggetto a *fluttuazioni* statistiche, le deviazioni casuali dal comportamento deterministico lungamente discusse nel capitolo IV. Poiché non abbiamo informazioni sufficienti che ci permettono di scrivere un'equazione

principale per la sottostante distribuzione di probabilità, assimiliamo l'effetto delle fluttuazioni ad una *forza casuale*, $F(t)$, come discusso alla fine del capitolo IV, § 2. L'equazione del bilancio dell'energia 6.7 diventa ora un'equazione stocastica differenziale della forma

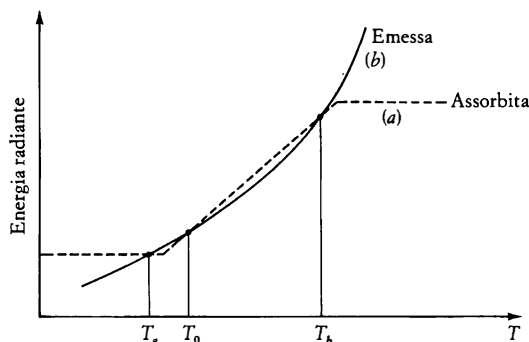
$$\frac{dT}{dt} = -\frac{\partial U(T)}{\partial T} + F(t) \tag{6.8}$$

L'importante nuovo elemento introdotto da questa descrizione allargata è che stati differenti diventano *connessi* per mezzo delle fluttuazioni. Detto diversamente, partendo da un qualche stato iniziale, il sistema raggiungerà prima o poi, ogni altro stato. Questo rimane vero anche per gli stati stabili T_a e T_b, la cui stabilità deterministica viene trasformata in un qualche tipo di *metastabilità*. Come discusso nel capitolo IV, § 5, la scala dei tempi di questo fenomeno è determinata da due fattori: la *barriera di potenziale*

$$\Delta U_{a,b} = U(T_0) - U(T_{a,b})$$

e la forza delle fluttuazioni come viene tipicamente misurata dalla loro *varianza*, q^2, di $F(t)$. L'ordine di grandezza della stima già dato nel capitolo IV, § 5, equazione 4.31, porta a:

Figura 6.5

Bilancio globale dell'energia del pianeta Terra. Le curve (a) e (b) descrivono, rispettivamente, la dipendenza dalla temperatura della parte assorbita dell'energia solare in arrivo e della radiazione infrarossa emessa nello spazio dalla terra. Queste dipendenze permettono l'esistenza di tre regimi di stati stazionari, due dei quali (T_a, T_b) sono stabili e corrispondono, rispettivamente, a un clima glaciale e a un clima simile a quello attuale.

Tempo medio di transizione dallo stato T_a allo stato T_b tramite lo stato instabile T_0

$$\tau_{a,b} \sim \exp\left(\frac{\Delta U_{a,b}}{q^2}\right) \tag{6.9}$$

È chiaro che se le fluttuazioni sono ragionevolmente piccole questo tempo sarà eccessivamente grande per ogni grandezza della barriera di potenziale. Valori tipici risultano essere da 10^4 a 10^5 anni, ovvero, l'intervallo di tempo caratteristico della glaciazione quaternaria. Non si può tuttavia pensare che l'equazione 6.9 fornisca una spiegazione dei cicli di glaciazione, poiché il passaggio fra due stati climatici T_a e T_b rimane un processo casuale, senza una ben definita periodicità. È qui che l'influenza di una debole forzatura periodica esterna diventa cruciale.

Ripetiamo l'analisi di cui sopra aggiungendo alla costante solare Q un termine della forma $Q\varepsilon$ sen ωt, il cui periodo $\Delta = 2\pi/\omega$ rappresenta l'effetto delle variazioni orbitali o, in effetti, di ogni variazione che influenza la quantità di energia solare ricevuta dalla Terra. Ne risulta che, se Δ è molto piú piccolo del tempo di transizione $\tau_{a,b}$ la forzatura esterna non ha praticamente effetto. Ma quando Δ diventa paragonabile a $\tau_{a,b}$ la risposta del sistema alla forzatura viene accresciuta enormemente. Presumibilmente, la presenza della forzatura abbassa la barriera di potenziale e facilita in questo modo il passaggio fra stati climatici. Inoltre, a causa della periodicità del segnale, il sistema viene trasportato e chiuso in un passaggio ritmico fra due tipi di clima attraverso la barriera cosí abbassata. Cominciamo dunque a vedere un meccanismo qualitativo di glaciazioni in cui sia i processi generati internamente sia i fattori esterni giocano un ruolo importante. Il semplice modello discusso sopra permette anche di identificare i parametri che piú probabilmente influenzano il comportamento del sistema.

Naturalmente, il carattere schematico del modello richiama lo sviluppo di una piú sofisticata descrizione che, mentre deve ancora essere riconducibile ad uno studio analitico, incorpora addizionali variabili chiave nella descrizione. Una classe particolarmente interessante di tali modelli è quella che descrive le interazioni fra la temperatura media superficiale e il ghiaccio continentale o marino. Per valori plausibili dei parametri essi predicono soluzioni periodiche nel tempo del tipo ciclo-limite, il cui periodo è anche paragonabile ai tempi caratteristici della glaciazione. Tuttavia, incorporando nella descrizione l'effetto delle fluttuazioni, si mostra che ogni traccia del comportamento oscillatorio scomparirà dopo un sufficiente lasso di tempo, a causa delle proprietà

di stabilità debole della fase dell'oscillatore. Qui, nuovamente, l'accoppiamento con una forzatura esterna periodica permette la stabilizzazione della fase e dunque un regime di trascinamento e di accordo della fase fra il sistema climatico e la forzatura esterna.

Ulteriori progressi nell'entusiasmante campo della dinamica climatica dovrebbero assottigliare il divario fra semplici modelli qualitativi e dettagliate descrizioni numeriche. Ci si aspetta che un riesame dei dati climatici alla luce della teoria dei sistemi dinamici fornisca ulteriori delucidazioni, dalle quali dovrebbe emergere una nuova descrizione della complessità del nostro ambiente nella sua vastità. Si riassumeranno alcuni tentativi in questa direzione nell'*Appendice* D.

5. *Comportamento probabilistico e strategie di adattamento negli insetti sociali.*

Ci rivolgiamo ora verso un nuovo aspetto della complessità. Ci interesseremo ai fenomeni in cui le variabili di stato tradizionalmente utilizzate nelle scienze fisiche, come la temperatura o le concentrazioni chimiche, non sono piú le variabili nei cui termini si manifesta il comportamento complesso. Entrano in gioco nuovi elementi che richiedono un linguaggio in cui strategia, anticipazione, simbolica ritualizzazione diventano le parole chiave. Tuttavia, la possibilità di eseguire transizioni fra diversi modi di comportamento rimane la principale impronta della complessità. Ciò ci permette di applicare all'analisi i concetti e i metodi dei sistemi dinamici non-lineari in condizioni lontane dall'equilibrio. Cominciamo la nostra discussione con qualche aspetto del comportamento degli insetti sociali.

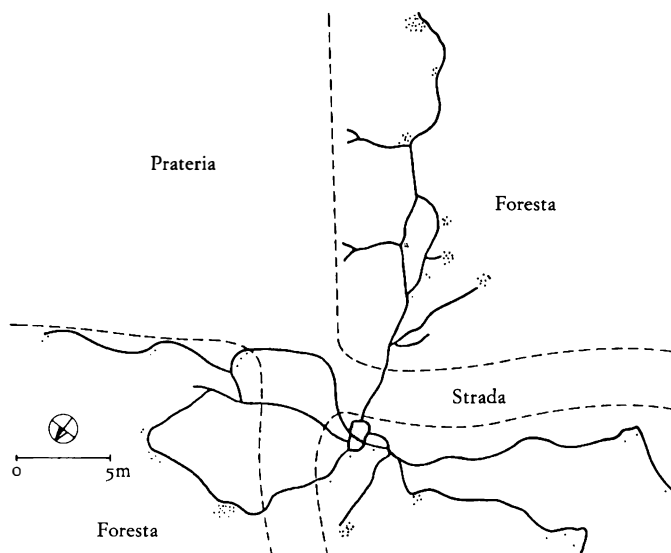
Gli insetti sociali (api, formiche, termiti), hanno segnato un enorme successo ecologico nell'evoluzione biologica. Hanno conquistato la maggior parte delle nicchie note, e in particolare i tropici, dove la loro biomassa può valere alcune centinaia di chilogrammi per acro.

Il punto di vista sulle società degli insetti sostiene che la costruzione del nido, la ricerca del cibo, e altre attività collettive sono i prototipi di un mondo deterministico in cui gli insetti individuali sono piccoli, fidati automi che obbediscono ad un programma genetico strettamente stabilito. Oggi, questo quadro di assoluta rigidità va sbiadendo, e un nuovo paradigma sta gradualmente emergendo in cui giocano un ruolo importante gli elementi casuali dall'ambiente e una plasticità del comportamento individuale.

Ciò che colpisce in una società di insetti è l'esistenza di due scale: una a livello dell'individuo, caratterizzata da un comportamento altamente probabilistico, e un'altra a livello della società come un tutto, dove, nonostante l'inefficienza e l'imprevedibilità degli individui, si sviluppano al livello dell'intera colonia, modelli coerenti caratteristici della specie. Vediamo come questi due aspetti sono legati per assicurare una organizzazione globale della società. Scegliamo come primo esempio la ricerca del cibo delle formiche, e confrontiamo il corso di questa attività in due diverse situazioni: nella prima esiste nelle vicinanze del nido una sorgente prevedibile di cibo, come una colonia di insetti di specie piú piccola (per esempio afidi), la cui vita media è dell'ordine di parecchi mesi; nella seconda si ha l'improvvisa disponibilità di una sorgente non prevedibile, come un uccello morto.

Nel primo caso si osserva la formazione di strade stabili dal nido alla colonia vicina, e ognuna di esse ha i suoi utenti specializzati (cfr. figura 6.6). Inoltre, poche formiche si avventurano al di fuori di questi assi. Chiaramente, nelle circostanze esistenti, lo sviluppo di strutture per-

Figura 6.6

Rappresentazione schematica del nido e della rete di tracce di una società di *Lasius fuliginosus*: ⁓ : sentiero; ○: nido; ⁘: alberi con afidi.

manenti stabili con un basso livello di rumore, è un beneficio per la società.

All'opposto, la struttura permanente in un ambiente imprevedibile può compromettere la plasticità della colonia e portarla ad un regime subottimale. Una possibile reazione verso un tale ambiente è dunque quella di mantenere un alto tasso di esplorazione e la capacità di sviluppare rapidamente strutture temporanee adatte per esplorare ogni occasione favorevole che possa nascere. In altre parole, sembra che la casualità presenti un valore di adattamento nell'organizzazione della società. Questa affermazione radicale sembra essere sostenuta dai dati sperimentali, ai quali rivolgiamo la nostra attenzione.

La scoperta di una nuova sorgente di cibo richiede la mobilitazione di formiche per assicurare la sua efficiente esplorazione. I meccanismi con cui le società delle formiche riescono a riunire un grande numero di individui attorno alle sorgenti di cibo costituiscono il *reclutamento* per il cibo. Studiamo il reclutamento per il cibo in due specie di formiche, *Tetramorium impurum* e *Tapinoma erraticum*. La prima specie può eseguire un reclutamento di gruppo, in altre parole, la formica reclutatrice traccia un sentiero mentre ritorna al nido dalla fonte di cibo e successivamente conduce un piccolo numero di reclute lungo il sentiero verso la fonte di cibo. Entrambe le specie possono eseguire, d'altra parte, un reclutamento di massa, in cui la traccia è sufficiente per guidare le reclute che si muovono indipendentemente dalla reclutatrice.

Esponiamo ora entrambe le specie alle stesse condizioni sperimentali. In particolare, si inizia il reclutamento per il cibo offrendo 1 mole di soluzione di saccarosio a 10 cm dal nido dopo 4 giorni di digiuno. Il comportamento delle reclute e delle reclutatrici viene videoregistrato. L'accuratezza del reclutamento viene misurata dalla proporzione di formiche reclutate dalla prima reclutatrice che ha effettivamente raggiunto la fonte di cibo, e anche dalla lunghezza media di una singola traccia di reclutamento effettivamente seguita dalle reclutatrici che viaggiano sole. La tabella 6.1 riassume i risultati. È chiaro che la traccia di reclutamento è uno stimolo molto piú efficiente per la *Tapinoma* che per la *Tetramorium*. Inoltre, il reclutamento di massa è molto piú accurato nella *Tapinoma*. Sotto le stesse condizioni sperimentali, non meno del 73 delle reclute che lascia il nido raggiunge la fonte seguendo la traccia segnata da una singola reclutatrice. In altre parole, la quantità di errore durante il reclutamento per il cibo varia considerevolmente da specie a specie. Mostriamo ora che esso può variare anche entro la stessa specie, a seconda delle condizioni.

Due contenitori ($5 \times 20 \times 30$ cm^3) vengono collegati con un ponte. Il nido viene piazzato in un contenitore, due fonti di cibo nell'altro. Il reclutamento verso le soluzioni di saccarosio viene iniziato dopo 8 giorni di digiuno.

Le fonti di cibo sono distanti 11 cm dal ponte e 14 cm una dall'altra. Esse vengono depositate simultaneamente o a un'ora di distanza, e il transito delle formiche *Tetramorium* viene videoregistrato. Risulta che il reclutamento è scarsamente accurato. Solo dal 20 al 40% delle reclute partecipa alla raccolta di cibo. La maggior parte delle formiche perde la traccia del reclutamento ed esplora l'area del foraggiamento. Esse scoprono velocemente la seconda fonte, anche quando il reclutamento verso la prima fonte era già ben stabilito. Tipicamente, quando due fonti sono di uguale qualità, una viene sfruttata al massimo finché si esaurisce; poi la seconda fonte viene completamente colonizzata e il suo sfruttamento si intensifica, cosicché non si registra nessuna interruzione nel tasso di raccolta del cibo. Quando la seconda fonte è piú concentrata in glucosio della prima, le formiche spostano i loro sforzi collettivi verso la fonte piú remunerativa senza tuttavia abbandonare completamente la prima. Il comportamento ricorda fortemente la biforcazione di due stati simultaneamente stabili (cfr. figura 3.9) e la sua deformazione per effetto di una piccola imperfezione (cfr. figura 3.10). Inoltre, vediamo chiaramente come la casualità permette al sistema di spostarsi fra questi due modi di comportamento.

Poniamo ora la seguente domanda. Le formiche che perdono la traccia a causa di errori sono in grado di scoprire nuove sorgenti di cibo, ma non sfruttano quelle già note. Qual è il miglior bilancio fra le fluttuazioni, che permettono scoperte e innovazioni, e l'accurato determinismo, che permette un immediato sfruttamento? Il seguente semplice modello fornisce qualche chiave.

Tabella 6.1.

Confronto tra la precisione di reclutamento di due specie di formiche (i numeri in parentesi sono relativi alle formiche effettivamente osservate).

		Tetramorium impurum	Tapinoma erraticum
Lunghezza % delle singole tracce effettivamente seguite		17 (40)	67.7 (47)
Recluta %	sola	8.9 (45)	73.6 (216)
Raggiungimento del cibo	in gruppo	6.0 (10)	—
Sorgente	Totale	18.2 (55)	73.6 (216)

Descriviamo dapprima come agisce il reclutamento globale in presenza di una singola fonte di cibo. Sia X il numero delle operaie alla fonte di cibo. Il flusso medio di formiche che arrivano alla stessa fonte, J_+, si può ragionevolmente considerare proporzionale al numero di incontri fra gli X individui e gli $N-X$ rimanenti, dove N è il numero totale di formiche in grado di partecipare al reclutamento. Cosí:

$$J_+ = aX(N-X) \tag{6.10a}$$

dove a è il tasso di reclutamento per individuo. Il flusso di partenza dalla fonte di cibo, J_-, è, d'altra parte, unicamente

$$J_- = -bX \tag{6.10b}$$

dove b è l'inverso del tempo medio speso per stare vicino al cibo e tornare indietro al nido. Quando la fonte di cibo è esaurita, quest'ultimo termine controlla da solo l'evoluzione di X. Ma quando la fonte contiene ancora cibo, il tasso globale di cambiamento di X è dato da:

$$\frac{dX}{dt} = aX(N-X) - bX \qquad \text{(cibo non esaurito)} \tag{6.11}$$

Questa equazione è nota come *equazione logistica*, e viene largamente usata in una grande varietà di problemi sulle popolazioni. Il valore di soglia di N è detto capacità di trasporto. Se $X < N$, J_+ è positivo, e questo esprime il *feedback* positivo della popolazione X in se stessa benché con un rallentamento espresso dal fattore di saturazione $N-X$. Ma se $X > N$, il *feedback* di cui sopra sparisce poiché J_+ diventa negativo.

Supponiamo ora che la colonia venga confrontata con k fonti di cibo identiche, regolarmente distribuite che sono equidistanti dal nido. Il reclutamento viene iniziato verso la fonte mediana, e si assume che le formiche siano distribuite attorno alla sorgente secondo una distribuzione normale. Si suppone che le formiche che raggiungono una fonte inizino il reclutamento verso di essa.

Una diretta generalizzazione delle equazioni 6.10-6.11 porta a:

$$\left.\begin{aligned} &\frac{dX_i}{dt} = \sum_{j=1}^{k} a_{ji}X_j(N - X_1 - \cdots - X_k) - bX_i && (6.12a)\\ &\qquad\qquad \text{(fonte di cibo non esaurita)} \\ &\frac{dX_i}{dt} = -bX_i \quad \text{(fonte di cibo esaurita)} && (6.12b) \end{aligned}\right\} \quad i = 1, \ldots, k$$

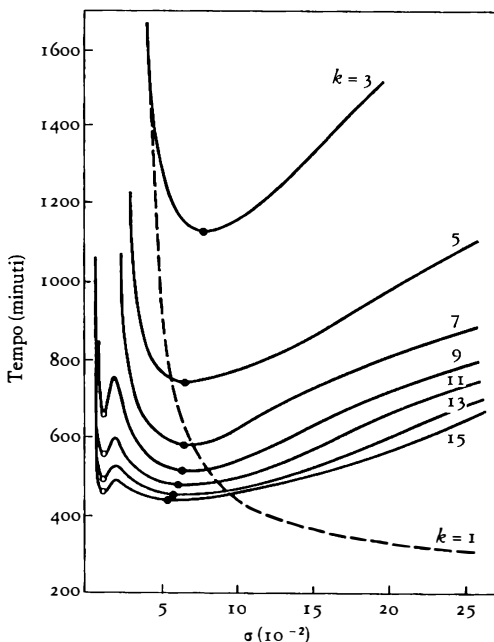
dove $X_1, \ldots, X_k$ rappresentano ora il numero di formiche reclute verso la fonte 1, ..., k. Il coefficiente a_{ij}, $j \neq i$ è il tasso di reclutamento delle formiche reclutate in j che raggiungono per errore la fonte i. Si può stimare come:

$$a_{ji} = \frac{1}{S} \int_{|r_i - r| \leqslant d} dr \exp\left(-\frac{(r_j - r)^2}{2\sigma^2}\right) \tag{6.13}$$

dove S è un opportuno fattore di normalizzazione, $2d$ la dimensione della fonte di cibo, e σ^2 la varianza della distribuzione di probabilità di Gauss. Piú piccolo è σ e piú accurato è il reclutamento.

La figura 6.7 descrive i risultati di una soluzione numerica in cui il tempo necessario per sfruttare pienamente una data quantità di cibo dipende dall'errore come viene espresso dalla varianza σ della distribuzione statistica, per dati numeri delle fonti di cibo. Come ci si aspettava,

Figura 6.7

Tempi di raccolta per numero di fonti di cibo k. Valori dei parametri $N = 650$, $a = 10^{-3}$, $b = 1,6 \times 10^{-3}$, $2d = 1,4$ cm.

quando è presente solo una sorgente la miglior strategia di reclutamento è quella che funziona senza errore. Tuttavia, quando il cibo viene distribuito c'è un valore «ottimale» dell'errore che minimizza il tempo di esplorazione. Questo definisce il livello di casualità durante la comunicazione che può essere vantaggiosa aumentando la possibilità di scoperte che permettono alla società di concentrare le sue attività sulle risorse piú remunerative e di promuovere la colonizzazione delle risorse che verranno completamente sfruttate in seguito.

Fino ad ora nella nostra discussione le singole formiche si differenziavano solo per il fatto che qualche aspetto del loro comportamento era distribuito statisticamente. C'è tuttavia, all'interno di una società di insetti, una divisione del lavoro. Un nido di formiche è una società femminile, in cui il ruolo dei maschi, che hanno vita breve, è limitato alla fecondazione. Questa divisione del lavoro fra le regine, che assicurano la riproduzione, e le operaie, che portano avanti i compiti necessari per il funzionamento della società, è descritta in molti libri sulla vita degli insetti.

In certe specie il gruppo delle operaie è composto di molte caste fisicamente distinte, ed è ben stabilita una divisione del lavoro basata sull'età. Sorprendentemente, tuttavia, sembra che entro una data casta ed entro un dato gruppo, le operaie manifestano notevoli differenze di attività il cui significato per la società non è chiaro.

Quando le formiche si trovano in condizioni sfavorevoli, esse trasferiscono i loro nidi. Il trasloco procede per esplorazione, scoperta del luogo favorevole, e trasporto del contenuto del nido originale. È stato progettato un esperimento in cui una colonia è stata oggetto di una serie di traslochi in un periodo di 8 mesi, ed è stato misurato il numero di carichi trasportati da ogni individuo. Un risultato molto sorprendente è che sembra esserci una certa stabilità nel livello di attività: le formiche risultate attive in un primo esperimento rimangono attive per parecchi mesi, e l'analogo accade per le formiche «pigre». Queste differenze sono non solo intrinseche agli individui, ma sembra siano modulate dall'ambiente e dalla storia del particolare individuo. Infatti, quando i due gruppi vengono separati, le formiche pigre diventano attive. Se i due gruppi si uniscono nuovamente, la colonia manterrà la memoria della «perturbazione» per un certo tempo.

In conclusione, una società di insetti sembra possedere una notevole plasticità che le permette di eseguire, con un minimo di programmazione a livello genetico, una serie di compiti complessi. Ancora una volta, le transizioni verso diversi modi di comportamento innescate dalle con-

dizioni ambientali sembrano essere il meccanismo primario alla base di questa straordinaria impresa.

6. *Autoorganizzazione nei sistemi umani.*

La nostra esperienza quotidiana ci insegna che l'adattabilità e la plasticità del comportamento, due fattori basilari dei sistemi dinamici non-lineari in grado di eseguire transizioni in condizioni lontane dall'equilibrio, vengono classificate fra le caratteristiche piú cospicue delle società umane. È dunque naturale aspettarsi che i modelli dinamici che permettono l'evoluzione e il cambiamento debbano essere i piú adeguati per i sistemi sociali.

Un modello dinamico della società umana inizia con il riconoscimento che, in aggiunta alla sua struttura interna, il sistema è fermamente inserito in un ambiente con il quale scambia materia, energia e informazione. Pensiamo, ad esempio, a una città in cui arrivano continuamente le materie prime e i prodotti agricoli, i prodotti finiti vengono esportati, mentre i mass-media mantengono i vari gruppi al corrente della situazione presente e delle tendenze immediate.

L'evoluzione di un tale sistema è un'interazione fra il comportamento dei suoi attori e i vincoli imposti dall'ambiente. È qui che il sistema umano trova la sua unica specificità. Contrariamente alle molecole, gli «attori» di un sistema fisico-chimico, o anche alle formiche o ai membri di ogni altra società animale, gli esseri umani sviluppano in effetti *progetti e desideri*. Alcuni di questi, infatti, provengono dalle anticipazioni su come possa ragionevolmente essere il futuro e dalle supposizioni concernenti i desideri degli altri attori. La differenza fra il comportamento desiderato e quello effettivo agisce dunque come un vincolo di nuovo tipo che, insieme con l'ambiente, forma la dinamica. Una questione fondamentale che si può sollevare è se, in queste circostanze, l'evoluzione nel suo insieme è in grado di portare verso un qualche tipo di condizione ideale globale, o, al contrario, se ogni sistema umano costituisce una realizzazione unica di un processo stocastico complesso le cui regole non possono in alcun modo essere progettate in anticipo. In altre parole, è sufficiente l'esperienza passata per prevedere il futuro, o, al contrario, non è l'alto grado di impredicibilità del futuro l'essenza dell'avventura umana, sia essa a livello individuale di apprendimento, o a livello collettivo di realizzazione della storia? Gli sviluppi esposti nei capitoli precedenti suggeriscono che la risposta a questa questione

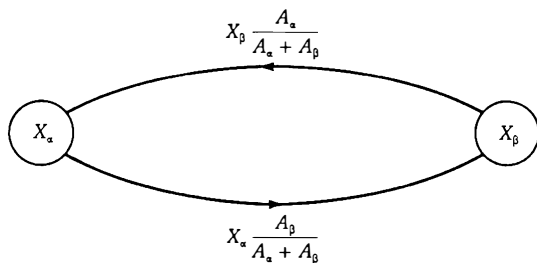
dovrebbe pendere verso la seconda alternativa. Vediamo se i modelli matematici ci permettono di giustificare questa sensazione intuitiva e, allo stesso tempo, di specificare piú nettamente la natura dell'imprevedibilità del sistema umano.

Cerchiamo dapprima di stimare il guadagno associato a una certa scelta. Possiamo supporre che il desiderio di prendere una particolare decisione i fra le K alternative, per unità di tempo, sia proporzionale alla attrazione relativa. Tuttavia, quando viene adottata una data opzione, la conclusione in generale cambierà, e cosí lo schema della scelta a livello della popolazione la rifletterà con il divenire delle scelte piú o meno attraenti. Questo *feedback* è mostrato nella figura 6.8 nel caso particolare di due scelte α e β. Qui A_α rappresenta l'attrazione della scelta α, X_α il numero di persone che hanno optato per α in un dato momento. Chiaramente il numero di individui che sperano di spostarsi sulla scelta β saranno quelli che hanno optato qualche altra scelta come α moltiplicata per la relativa attrazione di β, $A_\beta/(A_\alpha + A_\beta)$. Analogamente quegli individui che sperano di lasciare la scelta β in favore della scelta α saranno uguali a X_β moltiplicato per la relativa attrazione di α, $A_\alpha/(A_\alpha + A_\beta)$. Ne risulta un insieme di equazioni di equilibrio per X_α e X_β che appaiono molto simili alle equazioni ecologiche 6.11 e 6.12 ottenute nel precedente capitolo.

$$\frac{dX_\alpha}{dt} = aX_\alpha\left(\frac{X_\beta A_\alpha}{A_\alpha + A_\beta} - \frac{X_\alpha A_\beta}{A_\alpha + A_\beta}\right)$$

o, ricordando che $X_\beta = N - X_\alpha$, N essendo la grandezza della popolazione totale:

Figura 6.8

Ciclo di *feedback* che descrive la scelta sotto l'influenza di due opzioni simultaneamente disponibili α e β le cui attrazioni sono A_α e A_β, X_α, X_β popolazioni che hanno scelto α o β a un dato momento.

$$(6.14) \qquad \frac{dX_\alpha}{dt} = aX_\alpha\left(\frac{NA_\alpha}{A_\alpha + A_\beta} - X_\alpha\right)$$

e una simile equazione vale per X_β. Confrontando con l'equazione 6.11, vediamo che l'esistenza delle opzioni influisce la capacità di trasporto del sistema e lo rende una funzione dello stato istantaneo, attraverso la dipendenza dell'attrazione sugli X.

La precedente discussione si può facilmente estendere a un numero arbitrario di scelte K, e alla piú realistica situazione in cui l'attrazione di un'opzione i può ben dipendere dalla particolare popolazione j che intravede questa opzione. Si arriva cosí al seguente insieme di equazioni:

$$(6.15) \qquad \frac{dX_i}{dt} = CX_i\left[1 - \frac{X_i}{\sum_{j=1}^{K} \frac{N_j A_{ij}}{\sum_l A_{lj}}}\right] \qquad (i = 1, \ldots, K)$$

dove si assume ulteriormente che la popolazione totale è disomogenea e si rompe cosí in sottopopolazioni distinguibili N_j, ciascuna con il suo proprio punto di vista circa la relativa attrazione delle scelte.

Il comportamento del sistema dinamico descritto dalle equazioni 6.15 dipende in modo cruciale dal particolare modo in cui le attrazioni e le N_j dipenderanno dalle variabili della popolazione X_i, che caratterizzano lo stato istantaneo. Questo a sua volta dipende dal particolare sistema in considerazione. Nel seguito discutiamo questo problema in connessione con l'evoluzione delle strutture urbane.

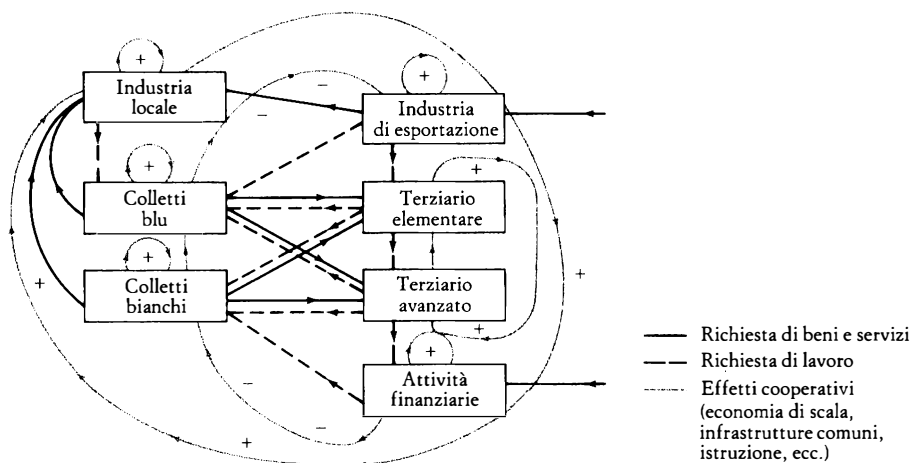
Consideriamo lo sviluppo di un centro urbano, con certe aree che si specializzano in specifiche attività economiche e con la segregazione residenziale che produce quartieri che differiscono nelle condizioni di vita e di accesso a lavori e servizi. Si riconoscono due componenti principali della popolazione, che corrispondono essenzialmente agli operai e agli impiegati. Per quanto riguarda i tipi di lavoro, essi possono appartenere all'industria (di esportazione o locale), alle funzioni terziarie (servizi fondamentali e specializzati), e le attività finanziarie. Le variabili principali che determinano lo stato sono ora il numero di residenti X_i^k del gruppo $k (k = 1, 2)$ al punto i, e i corrispondenti lavori J_i^k. I punti i, che indicano le diverse «opzioni» offerte agli individui, sono pensati come i siti di un reticolo che rappresenta lo spazio a disposizione. I

meccanismi di interazione fra le diverse variabili sono riassunti nella figura 6.9.

Per scrivere la forma modificata delle equazioni 6.15 per le X_i^k, si devono esprimere le quantità N_i e A_{ij} in termini di J_i^k e X_i^k. Si può ragionevolmente assumere che la sottopopolazione N_i sia essenzialmente determinata dal numero di lavori del tipo k situato al punto i. Per quanto riguarda le attrazioni A_{ij}, esse sono funzioni complicate di X_i^k, J_i^k e della distanza che separa i punti i e j. Non è necessario specificare qui la loro forma esplicita, ma mettiamo in evidenza che essa comprende fattori quali il costo e il tempo per andare al lavoro, il prezzo della terra, il carattere dei vicini, ecc.

Si possono scrivere equazioni analoghe per le altre categorie di attori, che descrivono, per esempio, la necessità per l'occupazione industriale di essere situata in un luogo con un comodo accesso dall'esterno e, preferibilmente, in un posto adiacente a industrie ivi già stabilite.

Il modello ora definito mostra l'evoluzione come un processo autonomo il cui corso è determinato a ogni momento dal meccanismo di interazione fra diversi attori. I vincoli dell'ambiente agiscono attraverso i parametri, e le condizioni iniziali adottate nel processo di soluzione esprimono l'effetto di casualità o di un sistematico intervento esterno o

Figura 6.9
Rappresentazione schematica delle interazioni e dei flussi di informazioni in un modello semplificato di una città.

«progettazione». Uno scenario alternativo, piú vicino a ciò che avviene nella realtà, è di lasciare evolvere il sistema per un certo periodo di tempo, modificare brutalmente il suo stato inserendo una nuova attività o una «innovazione», lasciare di nuovo che il sistema segua la sua dinamica autonoma finché viene inserita una nuova innovazione, e cosí via. Poiché le equazioni sono fortemente non-lineari, ci si aspetta che vi saranno parecchi rami di soluzioni che esibiscono un complicato insieme di fenomeni di biforcazione. Differenti condizioni iniziali situeranno il sistema in diversi bacini di attrazione, passando cosí a diverse evoluzioni, diverse «storie». La registrazione di una particolare storia fra la molteplicità di quelle possibili, non riflette certamente l'azione di un «progettatore globale» che tenta di ottimizzare qualche funzione globale, ma semplicemente che questo particolare schema è un modo di comportamento stabile e vitale. Abbiamo dunque sufficienti elementi a nostra disposizione per rispondere alla questione sollevata all'inizio del paragrafo.

Le soluzioni numeriche delle equazioni del modello stabiliscono inequivocabilmente l'esistenza di un grande numero di soluzioni e di intricati fenomeni di biforcazione. Partendo da uno spazio in cui le case sono inizialmente distribuite a caso, si osserva gradualmente la nascita di un disegno organizzato con i suoi centri amministrativi e di affari, le sue zone industriali, i centri di vendita, e i quartieri residenziali di varia qualità. In assenza di un qualche disturbo imponente, la struttura rimane indefinitamente stabile. Questa spontanea rottura di simmetria è molto simile alla formazione di strutture in cinetica chimica e in idrodinamica viste nel capitolo I, §§ 3 e 4.

Un risultato molto interessante che scaturisce dal modello è il seguente. Se una nuova attività viene inserita in un certo momento, essa crescerà e si stabilizzerà. Se il posto è ben scelto, esso può anche impedire la riuscita di analoghi tentativi fatti nelle vicinanze in un momento successivo. Tuttavia, se la stessa attività viene inserita piú tardi essa può non avere successo: regredirà a zero rappresentando una totale perdita. Questo illustra i pericoli del breve termine, della scarsa progettazione basata sulla diretta estrapolazione dell'esperienza passata. Tali metodi statici rischiano di fossilizzare la società o, in tempi lunghi, di farla crollare. Il messaggio della modellizzazione dinamica sostenuto in questo paragrafo è, al contrario, che la possibilità di adattamento delle società sia la principale fonte che permette loro di sopravvivere sul lungo periodo, di rinnovarsi, e di produrre originalità.

Appendici

A

Analisi della stabilità lineare

1. *Equazioni fondamentali.*

Nel capitolo II, §§ 6 e 7, sono stati definiti i concetti di stabilità e di biforcazione, mentre nel capitolo III sono state discusse numerose proprietà generali della ramificazione delle soluzioni e la loro stabilità. In questa appendice, rivolta al lettore piú sofisticato matematicamente, si delinea il formalismo che permette il calcolo della stabilità delle soluzioni di un sistema dinamico. Nell'*Appendice* B questo formalismo viene completato da un accenno al calcolo delle soluzioni che biforcano al di là della stabilità.

Abbiamo ripetutamente sottolineato che la proprietà della stabilità si riferisce alla risposta del sistema alle perturbazioni di vario tipo. Immaginiamo dunque un certo «stato di riferimento», X_{1s}, ..., X_{is}, ..., dove $\{X_i\}$ rappresenta l'insieme delle variabili di stato, che è continuamente sondato da fluttuazioni interne o disturbi esterni. Se trattiamo con un sistema completamente non distorto, libero da forze esterne e soggetto a un ambiente completamente uniforme e indipendente dal tempo, $\{X_{is}\}$ rappresenta ovviamente la soluzione indipendente dal tempo e spazialmente uniforme che ci si aspetta esista sotto queste condizioni, e che descrive l'assenza di autoorganizzazione e di comportamento complesso entro il sistema. Piú generalmente, possiamo scegliere che $\{X_{is}\}$ sia la soluzione che possiede le piú alte simmetrie permesse dai vincoli esterni. Nella terminologia del capitolo II, § 7, $\{X_{is}\}$ risiede sul *ramo termodinamico*, la continuazione del comportamento vicino all'equilibrio all'aumentare della distanza dall'equilibrio. L'autoorganizzazione e l'inizio del comportamento complesso si può dunque vedere come una transizione da $\{X_{is}\}$ a soluzioni di nuovo tipo.

Costruiamo ora un appropriato formalismo che ci permetta di sondare la stabilità dello stato di riferimento. La maggior difficoltà della teoria della stabilità risiede nella presenza di un gran numero di variabili che caratterizzano un problema, e nel fatto che in molti casi queste variabili sono funzioni delle coordinate spaziali. Introduciamo pertanto

una notazione piú sintetica, nella quale l'intero insieme di variabili è rappresentato da un vettore colonna, X, le cui componenti sono X_{1s}, ..., X_{is}, ... In una situazione indipendente dallo spazio, il numero totale di variabili, n (ad es. le concentrazioni chimiche), è tipicamente finito. X sarà allora definito nell'usuale spazio vettoriale familiare dall'algebra. Ma se alcune delle variabili dipendono dalla coordinata spaziale $\boldsymbol{r}$, oltre al precedente carattere vettoriale, X appartiene anche a uno spazio funzionale, come ad esempio, lo spazio delle funzioni il cui quadrato è integrabile, familiare dai corsi elementari di meccanica quantistica.

Con l'uso di questa notazione, possiamo esprimere nella forma seguente la velocità di cambiamento di X [cfr. equazioni (2.12) oppure (3.11)]:

$$\frac{\partial X}{\partial t} = F(X, \lambda) \tag{A.1}$$

In questo caso F è un *operatore* che agisce sullo spazio nel quale X è definito. Non è generalmente lineare, come risultato dei *feedback* che agiscono sul sistema. λ denota invece un insieme di parametri di controllo che influenzano l'evoluzione, come ad esempio, costanti di velocità, coefficienti di diffusione, ecc.

Lo stato di riferimento X_s precedentemente introdotto è esso stesso una particolare soluzione dell'equazione (A.1), quindi

$$\frac{\partial X_s}{\partial t} = F(X_s, \lambda) \tag{A.2}$$

Il ruolo delle perturbazioni sulla stabilità si esprimerà ponendo

$$X = X_s + x \tag{A.3}$$

dove x rappresenta il disturbo, e trasformando la A.1 in una equazione rispetto alla x:

$$\frac{\partial x}{\partial t} = F(X_s + x, \lambda) - F(X_s, \lambda) \tag{A.4}$$

Alcuni esempi aiuteranno il lettore a comprendere in che modo debbono essere intesi x e l'operatore F. Il caso piú semplice è certamente un problema che coinvolge una singola variabile, indipendente dallo

spazio, come il modello dell'equazione 2.15 in un mezzo ben rimescolato. La $\boldsymbol{F}$ si riduce allora allo scalare:

$$F = \text{velocità di reazione} = -k_2X^3 + k_1aX^2 - k_3X + k_4b$$

mentre il termine a destra della A.4 diviene:

$$\begin{aligned} F(X_s + x, \lambda) - F(X_s, \lambda) \\ &= -k_2(X_s + x)^3 + k_1a(X_s + x)^2 - k_3(X_s + x) + k_4b \\ &\quad -(-k_2X_s^3 + k_1aX_s^2 - k_3X_s + k_4b) \end{aligned}$$

ovvero, dopo aver sviluppato i binomi cubico e quadratico.

$$\text{(A.5)} \qquad \frac{\partial x}{\partial t} = (-3k_2X_s^2 + 2k_1aX_s - k_3)x + (-3k_2X_s + k_1a)x^2 - k_2x^3$$

Se, d'altra parte, prendiamo il modello del brussellatore in un mezzo ben rimescolato, equazione 3.18, otteniamo la seguente rappresentazione vettoriale a due componenti delle variabili di stato e delle funzioni di velocità:

$$\boldsymbol{X} = \begin{pmatrix} X \\ Y \end{pmatrix}, \qquad \boldsymbol{X}_s = \begin{pmatrix} A \\ B/A \end{pmatrix}, \qquad \boldsymbol{F} = \begin{pmatrix} A - BX + X^2Y - X \\ BX - X^2Y \end{pmatrix}$$

Un po' d'algebra di routine porta alla seguente forma esplicita delle equazioni (A.4):

$$\text{(A.6}a) \qquad \frac{\partial}{\partial t}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{bmatrix} (B-1)x + A^2y + \frac{B}{A}x^2 + 2Axy + x^2y \\ -Bx - A^2y - \frac{B}{A}x^2 - 2Axy - x^2y \end{bmatrix}$$

Se le reazioni avvengono in un mezzo non rimescolato, $\boldsymbol{F}$ comprenderà un addizionale contributo proveniente dal trasporto di materia attraverso la diffusione o la convezione. Per la particolare classe dei sistemi reazione-diffusione, le cui proprietà sono state a lungo descritte nel capitolo 1, § 4, si deve prendere in considerazione solo la diffusione. Come sottolineato nel capitolo 1, § 2, la legge di Fick, equazione 2.7*b*, offre spesso una eccellente rappresentazione di questo processo. Il cor-

rispondente operatore $\boldsymbol{F}$, ad esempio per il brussellatore, sarà allora della forma:

$$F_{dif}(X) = \begin{pmatrix} D_1 \nabla^2 X \\ D_2 \nabla^2 Y \end{pmatrix}$$

Per la dipendenza lineare da x e da y questo porta alla

$$\text{(A.6}b\text{)} \qquad \frac{\partial}{\partial t}\begin{pmatrix} x \\ y \end{pmatrix}_{dif} = \begin{pmatrix} D_1 \nabla^2 x \\ D_2 \nabla^2 y \end{pmatrix}$$

Si noti che F_{dif} opera sia sullo spazio dei vettori due-dimensionali sia sullo spazio delle funzioni dipendenti dalla coordinata $\boldsymbol{r}$.

Torniamo ora al formalismo generale, equazione A.4. Come suggerito dagli esempi espliciti trattati sopra, è naturale sviluppare in serie di potenze la parte destra attorno allo stato di riferimento X_s. Se $\boldsymbol{F}$ ha una struttura polinomiale rispetto a X questo è sempre possibile, e ci porta a un numero finito di termini. In situazioni piú intricate, comunque, $\boldsymbol{F}$ può dipendere da X in maniera diversa. Per affrontare il problema in questo caso assumiamo:

1. che $\boldsymbol{F}$ possa essere ancora sviluppata in serie di potenze rispetto a $\boldsymbol{x}$, e
2. che questo sviluppo possa essere troncato a un ordine finito.

Quest'ultima ipotesi ci limita allo studio della *stabilità infinitesima*, ovvero la ripondenza del sistema a piccoli disturbi tali che $|x|/|X_s| \ll 1$. Questo è lo stesso molto utile, poiché la stabilità infinitesima dà condizioni necessarie per la instabilità, nel senso che, se X_s è instabile per $\boldsymbol{x}$ piccoli, sarà instabile per ogni $\boldsymbol{x}$.

Formalmente lo sviluppo sopra descritto può essere rappresentato come:

$$\text{(A.7)} \quad F(X_s + x, \lambda) = F(X_s) + \left(\frac{\delta F}{\delta X}\right)_{X_s} \cdot x + \frac{1}{2}\left(\frac{\delta^2 F}{\delta X\, \delta X}\right)_{X_s} \cdot xx + \frac{1}{6}\left(\frac{\delta^3 F}{\delta X\, \delta X\, \delta X}\right)_{X_s} \cdot xxx + \cdots$$

Gli oggetti $(\delta F/\delta X)_{X_s}$, ecc. sono generalizzazioni del concetto di derivata, e nella letteratura matematica sono chiamate derivate di Fré-

chet. Per esempio, un paragone con le equazioni A.6*a* - A.6*b* ci porta a identificare nel seguente modo la prima derivata di Fréchet per il modello del brussellatore:

$$\left(\frac{\delta F}{\delta X}\right)_{X_s} = \begin{pmatrix} B - 1 + D_1 \nabla^2 & A^2 \\ -B & -A^2 + D_2 \nabla^2 \end{pmatrix} \tag{A.8}$$

questo mostra che $(\delta F/\delta X)_{X_s}$ opera come una matrice nello spazio dei vettori due-dimensionali e come un operatore differenziale nello spazio delle funzioni dipendenti dallo spazio.

2. *Il «principio» della stabilità linearizzata.*

Le equazioni A.7 e A.4 costituiscono un problema altamente non-lineare il quale, di regola, è altrettanto intrattabile di quello originale (equazione A.1). A questo punto però si può invocare un risultato piú importante dell'analisi che permette un ulteriore progresso. Questo teorema, anche noto come il principio della stabilità linearizzata, confronta le proprietà di stabilità dei due seguenti problemi:

- il problema originale, totalmente non-lineare, e
- il problema linearizzato, nel quale si omettono i termini di ordine piú grande.

Introducendo la notazione ristretta

$$\begin{aligned} \mathcal{L}(\lambda) &= (\delta F/\delta X)_{X_s} \\ \boldsymbol{h}(\boldsymbol{x}, \lambda) &= \frac{1}{2}\left(\frac{\delta^2 F}{\delta X\, \delta X}\right)_{X_s} \cdot \boldsymbol{x}\boldsymbol{x} + \cdots \end{aligned} \tag{A.9}$$

per i contributi rispettivamente lineari e non-lineari, possiamo formulare questi due problemi come le soluzioni dei seguenti insiemi di equazioni:

$$\frac{\partial \boldsymbol{x}}{\partial t} = \mathcal{L}(\lambda)\cdot \boldsymbol{x} + \boldsymbol{h}(\boldsymbol{x}, \lambda) \qquad \text{(problema non-lineare)} \tag{A.10}$$

e

$$\frac{\partial \boldsymbol{x}}{\partial t} = \mathcal{L}(\lambda)\cdot \boldsymbol{x} \qquad \text{(problema ausiliario linearizzato)} \tag{A.11}$$

Si noti che entrambi i sistemi sono omogenei, essendo identicamente soddisfatti dalla funzione banale $\boldsymbol{x} = 0$.

Il teorema della stabilità linearizzata conviene ora ciò che segue:

1. Se la soluzione banale $\boldsymbol{x} = 0$ del problema linearizzato (A.11) è asintoticamente stabile, allora $\boldsymbol{x} = 0$ (o equivalentemente $\boldsymbol{X} = \boldsymbol{X}_s$) è una soluzione asintoticamente stabile del problema non-lineare, equazioni A.11 o A.1.
2. Se la soluzione banale $\boldsymbol{x} = 0$ del problema linearizzato A.11 è instabile, allora $\boldsymbol{x} = 0$ (o equivalentemente $\boldsymbol{X} = \boldsymbol{X}_s$) è una soluzione instabile del problema non-lineare, equazioni A.11 o A.1.

Non riproduciamo qui la prova, ma, alla luce delle osservazioni fatte nell'ultima parte del § 1, dovremmo avere già sviluppato un'idea intuitiva circa la ragionevolezza delle affermazioni di questo teorema. Il punto importante da tenere a mente è che, in virtú del teorema, la stabilità, una delle proprietà fondamentali dei sistemi dinamici, è ridotta a un problema lineare: un problema ben piú tradizionale e trattabile dell'analisi.

3. *L'equazione caratteristica.*

Per procedere ulteriormente sarà istruttivo concentrarsi su una classe particolare di *sistemi autonomi*, per i quali i vincoli sono indipendenti dal tempo, e supporre che lo stato di riferimento $\boldsymbol{X}_s$ sia una soluzione di stato stazionario. Ne segue che l'operatore di stabilità lineare $\mathcal{L}(\lambda)$ è indipendente dal tempo. In queste ipotesi, si può facilmente controllare che le equazioni A.11 ammettono soluzioni della forma:

$$\boldsymbol{x} = \boldsymbol{u} e^{\omega t} \tag{A.12}$$

In altre parole la dipendenza dal tempo della soluzione è contenuta interamente nella funzione esponenziale (si noti che ω è generalmente a valori complessi), mentre il vettore $\boldsymbol{u}$ tiene conto sia della struttura di $\boldsymbol{x}$ come vettore nello spazio delle variabili di stato sia della sua dipendenza dalla coordinata spaziale $\boldsymbol{r}$.

Sostituendo nella (A.11) e ricordando che $\mathcal{L}(\lambda)$ è indipendente dal tempo, vediamo che l'esponenziale si elimina da entrambe le parti e si ottiene la

$$\mathcal{L}(\lambda) \cdot \boldsymbol{u} = \omega \boldsymbol{u} \tag{A.13}$$

Questa relazione definisce un *problema agli autovalori*, un classico problema dell'analisi. Questi problemi sorgono frequentemente in meccanica quantistica o nei problemi di trasferimento di calore o di massa. Ciò che è notevole rilevare è che, indipendentemente dalle proprietà degli autovettori $\boldsymbol{u}$, il problema è risolto una volta determinati gli autovalori ω. Infatti, ricordiamo dalla A.12 che $\boldsymbol{x}$ varia nel tempo come

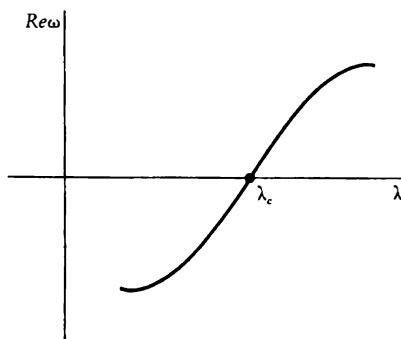
$$\boldsymbol{x} \sim e^{\omega t} = e^{(Re\omega)t} \cdot e^{i(Im\omega)t}$$

dove $Re\omega$ e $Im\omega$ sono, rispettivamente, le parti reale e immaginaria della quantità complessa ω. Ne segue che se $Re\omega < 0$, $\boldsymbol{x}$ è una funzione che decresce esponenzialmente (con o senza una modulazione oscillatoria a seconda che $Im\omega$ non sia o sia nulla). Quindi la soluzione di riferimento $\boldsymbol{x} = 0$ si raggiunge nel limite per $t \to \infty$, in altre parole, $\boldsymbol{x} = 0$ è asintoticamente stabile. Se, d'altra parte, $Re\omega > 0$, le perturbazioni crescono esponenzialmente, e lo stato di riferimento $\boldsymbol{x} = 0$ è instabile.

Questi due regimi, per i quali si applica il principio della stabilità linearizzata, sono separati da un regime noto come *stabilità marginale*, $Re\omega = 0$, il quale segnala la soglia di instabilità dello stato di riferimento.

Possiamo ora meglio capire l'importanza del concetto di parametro di controllo λ. Infatti una variazione di λ induce una variazione di $\mathcal{L}$ e, attraverso essa, dell'autovalore ω. L'esistenza di una transizione tra due regimi qualitativamente diversi sarà quindi riflessa dal fatto che almeno uno degli autovalori ω varierà come una funzione di λ nella forma descritta nella figura A.1 (cfr. anche la fig. 3.16). Il valore λ_c di λ per il

Figura A.1

Dipendenza della parte reale dell'esponente di stabilità ω dal parametro di controllo λ. Il valore $\lambda = \lambda_c$ segna la transizione dalla stabilità asintotica ($\lambda < \lambda_c$) alla instabilità ($\lambda > \lambda_c$).

quale Reω cambierà segno sarà il *valore critico* oltre il quale è impossibile il formarsi della instabilità.

La dipendenza specifica dell'autovalore ω dal parametro di controllo λ dipende dal problema considerato. Una classe molto importante di sistemi che si sono incontrati frequentemente nel corso dell'esposizione sono sistemi nei quali l'operatore $\boldsymbol{F}$ può essere decomposto in due parti, una esprimente la velocità $\boldsymbol{v}(\boldsymbol{x}, \lambda)$ di un processo locale come una reazione chimica, e un altro che esprime il trasporto di materia attraverso la diffusione. Una particolare illustrazione di questi sistemi di *reazione-diffusione* viene dal brussellatore (equazioni A.6*a* - A.6*b*).

Piú generalmente, per un sistema di reazione-diffusione a piú variabili, scriveremo

$$\boldsymbol{F} = \boldsymbol{v}(\boldsymbol{X}, \lambda) + \boldsymbol{D} \cdot \nabla^2 \boldsymbol{X} \tag{A.14}$$

il cui operatore linearizzato $\mathcal{L}$ è dato da

$$\mathcal{L} = (\partial \boldsymbol{v} / \partial \boldsymbol{X})_{X_s} + \boldsymbol{D} \cdot \nabla^2 \tag{A.15}$$

In molte situazioni le funzioni velocità, $\boldsymbol{v}$, sono indipendenti dallo spazio. Ne risulta che l'intera dipendenza spaziale delle soluzioni delle equazioni linearizzate (cfr. A.11)

$$\frac{\partial \boldsymbol{x}}{\partial t} = \left(\frac{\partial \boldsymbol{v}}{\partial \boldsymbol{X}}\right)_{X_s} \cdot \boldsymbol{x} + \boldsymbol{D} \cdot \nabla^2 \boldsymbol{x} \tag{A.16}$$

sarà condizionata dalle proprietà dell'operatore di Laplace, ∇^2. Il metodo classico per determinare le possibili forme di questa dipendenza spaziale, indipendentemente da tutte le complessità della cinetica che sono contenute nella funzione di velocità $\boldsymbol{v}$, è quello di risolvere il problema agli autovalori del laplaciano:

$$\nabla^2 \varphi_m(\boldsymbol{r}) = -k_m^2 \varphi_m \tag{A.17}$$

Le autofunzioni e gli autovalori $(\varphi_m, -k_m^2)$ dipenderanno naturalmente dalla geometria, dalle dimensioni e dalle condizioni al bordo. Per esempio, se si risolve l'equazione (A.17) su un segmento di lunghezza L con le condizioni al bordo dell'annullarsi di φ_m per $r = 0$ e $r = L$, si verifica facilmente che

$$k_m = \frac{m\pi}{L} \qquad \varphi_m(r) = \sin \frac{m\pi r}{L} \tag{A.18}$$

dove m è un qualunque intero non nullo. Se, d'altra parte, l'equazione (A.10) si risolve su uno strato bidimensionale circolare, φ_m sarà data dai prodotti delle funzioni di Bessel e delle funzioni trigonometriche.

Qualunque sia la forma dettagliata di φ_m, la soluzione dell'equazione lineare A.11 sarà (cfr. equazione A.12):

$$x = c\varphi_m(r)e^{\omega_m t} \tag{A.19}$$

dove c rende conto della differenza fra le variabili di stato, e l'autovalore ω dell'operatore linearizzato $\mathcal{L}$ è indicato dallo stesso indice m come l'autofunzione φ_m del laplaciano.

Sostituendo l'espressione (A.19) nell'equazione (A.16) vediamo che gli operatori $\partial/\partial t$ e ∇^2 agiscono dando, rispettivamente, ω_m e $-k_m^2$ moltiplicati per x stesso. Ne risulta che entrambi φ_m e $e^{\omega_m t}$, che appaiono come un fattore comune di tutte le componenti del vettore x, si cancellano. Il problema viene dunque ridotto a

$$\omega_m c = \left(\frac{\partial v}{\partial X}\right)_{X_s} \cdot c - Dk_m^2 \cdot c$$

ovvero, se D è diagonale

$$(\omega_m + D_i k_m^2)c_i = \sum_j \left(\frac{\partial v_i}{\partial X_j}\right)_{X_s} c_j$$

Questo è un insieme di equazioni algebriche omogenee per $\{c_i\}$. Esso ammette una soluzione non banale se il determinante dei coefficienti dei c_i si annulla. Questo dà luogo alla condizione di risolubilità, o *equazione caratteristica*:

$$\det \left| \left(\frac{\partial v_i}{\partial X_j}\right)_{X_s} - \delta_{ij}^{kr}(D_j k_m^2 + \omega_m) \right| = 0 \tag{A.20}$$

dove δ_{ij}^{kr} è il simbolo di Kronecher, $\delta_{ij}^{kr} = 0$ per $i \neq j$, $\delta_{ij}^{kr} = 1$ per $i = j$. Questa equazione determina gli autovalori ω_m in termini degli autovalori del laplaciano k_m^2, dei coefficienti di diffusione D_i, e degli altri parametri contenuti in $(\partial v_i/\partial X_j)_{X_s}$.

4. *Esempi.*

Illustriamo ora il tipo di informazione ottenuto dalla equazione A.20 attraverso alcuni esempi rappresentativi.

Una variabile. Un buon esempio è il modello dell'equazione 2.15, la cui equazione di velocità è data dalla A.5. In presenza della diffusione l'operatore linearizzato è la somma di uno scalare e del laplaciano:

$$\mathcal{L} = (-3k_2X_s^2 + 2k_1aX_s - k_3) + D\nabla^2$$

Ne risulta che l'equazione caratteristica si riduce alla

$$\omega_m = -3k_2X_s^2 + 2k_1aX_s - k_3 - Dk_m^2$$

e, piú generalmente per un sistema arbitrario a una variabile:

$$\omega_m = \left(\frac{\partial v}{\partial X}\right)_{X_s} - Dk_m^2$$

Segue quindi che ω_m è reale, e perciò la perturbazione attorno allo stato di riferimento evolve monotonamente nel tempo. Inoltre, il contributo della diffusione è negativo o zero; in altre parole, la diffusione non può destabilizzare questo sistema. La prima instabilità da X_s può venire pertanto solo dalla cinetica. Questo suggerisce che se esiste una nuova soluzione oltre l'instabilità, essa non mostrerà nessuna dipendenza intrinseca dallo spazio.

Due variabili. Un tipico esempio è il modello del brussellatore, per il quale la forma esplicita dell'operatore linearizzato $\mathcal{L}$ è data dall'equazione (A.8). L'equazione caratteristica corrispondente è data dall'annullarsi del determinante 2 × 2

$$\begin{vmatrix} B - 1 - (D_1k_m^2 + \omega_m) & A^2 \\ -B & -A^2 - (D_2k_m^2 + \omega_m) \end{vmatrix} = 0$$

o, in forma esplicita

$$\text{(A.22)} \qquad \begin{aligned} &\omega_m^2 - \omega_m[B - A^2 - 1 - (D_1 + D_2)k_m^2] + A^2 \\ &\qquad + [-(B-1)D_2k_m^2 + A^2D_1k_m^2] + D_1D_2k_m^4 = 0 \end{aligned}$$

Piú generalmente, in un sistema a due variabili l'equazione caratteristica sarà di secondo grado in ω_m e ammetterà pertanto eventuali soluzioni complesse coniugate. Ponendo $\omega_m = Re\omega_m + iIm\omega_m$, ricaviamo da

A.20 che $Re\omega_m$ si può annullare e $Im\omega_m$ rimanere finito se è soddisfatta la seguente condizione di stabilità marginale:

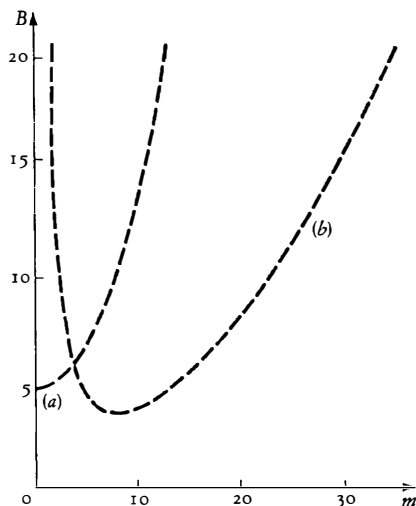
$$\left(\frac{\partial v_1}{\partial X_1}\right)_{X_s} + \left(\frac{\partial v_2}{\partial X_2}\right)_{X_s} = (D_1 + D_2)k_m^2 \qquad \text{(A.23}a\text{)}$$

Per il brussellatore questo ci dà

$$B \equiv B_m = A^2 + 1 + (D_1 + D_2)k_m^2$$

e fornisce una relazione tra un parametro; in questo caso è stato scelto B, e gli autovalori del laplaciano k_m^2. Per un sistema generale a due variabili un parametro di controllo λ, contenuto in $(\partial v_i/\partial X_j)_{X_s}$, può sempre essere scelto in modo che l'equazione A.23 sia rappresentata dalla curva (*a*) della figura A.2 (cfr. anche la figura 3.29). La corrispondente parte immaginaria può essere calcolata dalla A.20 nella quale si sostituisce (A.23*a*). Nel brussellatore questo ci dà

Figura A.2

Diagramma di stabilità lineare per il brussellatore in un mezzo monodimensionale. Il parametro di biforcazione B è disegnato in funzione del numero d'onda m. (*a*) mostra la regione di autovalori complessi, (*b*) la regione di autovalori reali. Valori dei parametri: $A = 2$; $L = 1$; $D_1 = 0{,}0016$; $D_2 = 0{,}008$.

$$\Omega_m^{\ 2} = (Im\,\omega_m)^2 = A^2 + [A^2 D_1 - (B_m - 1)D_2]k_m^{\ 2} + D_1 D_2 k_m^{\ 4}$$

Se potessero esistere nuove soluzioni oltre la curva di stabilità marginale, ci aspetteremmo su queste basi che esse abbiano un comportamento periodico con una *frequenza intrinseca* data essenzialmente da Ω_m. D'altra parte, esse avranno ancora una dipendenza banale dallo spazio, poiché il primo modo instabile che sarà eccitato corrisponde nuovamente a $k_m = 0$.

La situazione è molto differente se l'equazione A.20 ammette ω_m reali. La condizione di stabilità marginale diviene allora semplicemente $\omega_m 0$, oppure

$$\text{(A.23}b\text{)} \qquad \det\left|\left(\frac{\partial v_i}{\partial X_j}\right)_{X_s} - \delta_{ij}^{kr}\, D_i k_m^{\ 2}\right| = 0$$

Questa relazione è di quarto grado in k_m. Pertanto, se noi disegnamo un parametro di controllo λ [che è contenuto in $(\partial v_i/\partial X_j)_{X_s}$] in funzione di k_m ci possiamo aspettare un estremo in un certo k_{m_c} non nullo, come indicato nella curva (*b*) della figura 1.2 (cfr. anche la figura 3.29). Il primo modo instabile che sarà eccitato sarà pertanto dipendente dallo spazio. Il punto è che, alla luce dell'equazione A.23*b*, k_{m_c} è interamente determinato dai parametri del sistema, indipendentemente dalla grandezza, dalla geometria e perfino dalla dimensione. Inoltre, in accordo con le equazioni A.19 e A.17, k_{m_c} determina la lunghezza d'onda spaziale del disturbo che agisce su X_s, e per $\lambda \geqslant \lambda_c$ ci aspettiamo che questo disturbo determini le proprietà principali del sistema oltre l'instabilità. Abbiamo quindi un meccanismo per la genesi di una *lunghezza d'onda intrinseca* in un sistema fino ad ora spazialmente omogeneo. Chiameremo questa situazione *rottura della simmetria spaziale*.

5. *Sistemi che mostrano una dinamica caotica.*

Quando la dinamica diviene caotica, lo stato stazionario sul ramo termodinamico cessa di essere un appropriato stato di riferimento. Ciò nonostante, come menzionato brevemente nel capitolo III, § 8, è possibile caratterizzare l'instabilità del moto e la risultante casualità di comportamento grazie agli *esponenti di Ljapunov*. Si considerino due traiettorie dello spazio delle fasi con condizioni iniziali in due punti vicini, X_{0i} e $X_{0i} + \Delta x_{0i}$ rispettivamente ($i = 1, \ldots, n$), come mostrato in figura A.3. L'evoluzione del loro distacco $\Delta x_i \equiv \Delta x_i(x_0, t)$ nelle immediate vici-

nanze del punto X_{0i} è essenzialmente dato dal vettore tangente alla traiettoria che parte da X_{0i}. Può quindi essere trovata linearizzando le equazioni per le variabili X_i attorno ai loro valori istantanei $X_i(t)$. In accordo con quanto detto nel capitolo III, § 6 (cfr. equazione 3.25) questo ci dà

$$\frac{d\Delta x_i}{dt} = \sum_j L_{ij}(\{\Phi_m(t)\})\,\Delta x_j \qquad (i = 1, \ldots, n) \tag{A.24}$$

Introduciamo ora la velocità esponenziale media di divergenza di due traiettorie inizialmente vicine nel modo seguente:

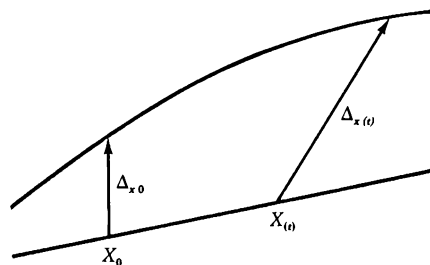
$$\lambda_L(X_0, \Delta x_0) = \lim_{\substack{t\to\infty \\ |\Delta x_0|\to 0}} \frac{1}{t} \ln \frac{|\Delta x(X_0, t)|}{|\Delta x_0|} \tag{A.25}$$

dove Δx denota il vettore dello spazio delle fasi le cui componenti sono $\{\Delta x_i\}$ e la cui norma è $|\Delta x|$. Si può mostrare che λ_L esiste ed è finito. Inoltre, poiché ci sono n direzioni indipendenti nello spazio delle fasi, λ_L assume n valori (eventualmente non distinti) che sono gli esponenti di Ljapunov. Questi vengono di solito ordinati per grandezza,

$$\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_n$$

e si può mostrare che sono indipendenti dalla metrica usata per calcolare la norma. Si noti che per ogni flusso generato da un sistema autonomo si deve annullare almeno un esponente di Ljapunov: nella direzione lungo il flusso $\Delta x(t)$ cresce solo linearmente, e questo ci dà nell'equazio-

Figura A.3

Definizione degli esponenti caratteristici di Ljapunov. Due condizioni iniziali si separano esponenzialmente all'evolvere del tempo.

ne (A.26) un $\lambda_L \sim (1/t) \ln t \sim 0$ per grandi valori di t. In un flusso caotico, d'altra parte, la divergenza esponenziale delle traiettorie si traduce con il fatto che almeno un esponente di Ljapunov è positivo, poiché $|\Delta \boldsymbol{x}(t)|/|\Delta \boldsymbol{x}_0 \sim \exp(\lambda_\alpha t)$ e quindi $(1/t) \ln |\Delta \boldsymbol{x}(t)|/|\Delta \boldsymbol{x}_0 \sim (1/t)\, \lambda_\alpha t \sim \lambda_\alpha > 0$.

Il concetto di esponente di Ljapunov può anche essere usato per descrivere la velocità media di espansione (o contrazione) $\lambda^{(p)}$ di un volume dello spazio delle fasi $\Delta V = \Delta x_{01} \cdots \Delta x_{0p}$, $p \leqslant N$. Per $p = n$ questo ci porta a

$$\lambda^{(n)} = \sum_{i=1}^{n} \lambda_i(X_0) \tag{A.26}$$

In un sistema conservativo il volume dello spazio delle fasi rimane invariato (cfr. capitolo III, § 2), e quindi $\lambda^{(n)} = 0$. In un sistema dissipativo, d'altra parte, la presenza di attrattori implica che ci dovrebbe essere una contrazione globale del volume dello spazio delle fasi, e quindi $\lambda^{(n)} < 0$. Pertanto la presenza di moti caotici in sistemi conservativi o dissipativi implica necessariamente che lo spazio delle fasi si sta espandendo lungo certe direzioni e contraendo lungo altre direzioni. È naturale concludere che in un moto caotico varietà e scelta sono generate in continuo lungo la direzione di espansione del moto, mentre lungo le direzioni che si contraggono si ottiene la prevedibilità. In altre parole, gli attrattori caotici (i quali, oltre alle precedenti proprietà, godono anche della asintotica stabilità) sono potenziali generatori ed elaboratori di informazione. Questo punto è stato discusso piú in particolare nel capitolo IV, § 8.

B

Analisi della biforcazione.

1. *Proprietà generali.*

L'importanza dell'analisi della stabilità lineare risiede nel fatto che essa mostra come un cambiamento qualitativo nel comportamento del sistema debba avvenire oltre il valore critico λ_c del parametro di controllo. Comunque, l'esistenza di nuove soluzioni fisicamente accettabili nascenti oltre la soglia dell'instabilità non potrà essere stabilita fintanto che non saranno state analizzate tutte le equazioni non-lineari. In questa appendice descriveremo brevemente alcuni passi base di questa analisi.

Il punto di partenza è dato dalle equazioni A.10,

$$\frac{\partial \boldsymbol{x}}{\partial t} = \mathcal{L}(\lambda)\cdot \boldsymbol{x} + \boldsymbol{h}(\boldsymbol{x}, \lambda) \qquad \text{(B.1)}$$

dove $\boldsymbol{h}$ rappresenta la parte non-lineare del lato destro della A.4. Prima di descrivere la costruzione delle soluzioni di biforcazione, sarà utile commentare alcune proprietà generali concernenti l'esistenza e la stabilità.

Nel capitolo III, §§ 4 e 5, abbiamo prodotto un numero di esempi (cfr. equazioni 3.10, 3.12 e 3.16) che mostrano come in certi casi è infatti possibile stabilire l'esistenza di nuovi rami di soluzioni emergenti dai valori del parametro nei quali lo stato di riferimento perde la sua stabilità. Ci si deve comunque trattenere dal dedurre, su queste basi, che sia sempre cosí.

Per convincerci che possono sorgere situazioni piú complicate, consideriamo il sistema non-lineare

$$\frac{dX_1}{dt} = X_1 + X_2^3 - \lambda X_1 \qquad \frac{dX_2}{dt} = X_2 - X_1^3 - \lambda X_2$$

Lo stato di riferimento è $X_{1s} = X_{2s} = 0$, e il problema linearizzato attor-

no a esso ha la forma (cfr. equazione A.20):

$$\omega x_1 = (1 - \lambda) x_1$$

$$\omega x_2 = (1 - \lambda) x_2$$

Questo sistema lineare di equazioni algebriche omogenee ha una soluzione non banale solo se $\omega = 1 - \lambda$. Per $\lambda = 1$ abbiamo uno stato di stabilità marginale. Effettivamente, nel modello in considerazione, $1 - \lambda$ è una radice doppia dell'equazione caratteristica, poiché nell'approssimazione lineare le due equazioni diventano non accoppiate. La questione che ora solleviamo è se al valore «critico» $\lambda = 1$ del parametro per il quale la soluzione dell'equazione caratteristica è $\omega = 0$, si osservano biforcazioni di nuovi rami di evoluzioni a stato stazionario. Per vedere questo moltiplichiamo la prima equazione per X_2, la seconda per X_1, e le sottraiamo. Si ottiene in questo modo: $X_1^4 + X_2^4 = 0$ che ovviamente non ha soluzioni reali (e quindi fisicamente accettabili) se non la soluzione banale. In altre parole, nonostante il fatto che il problema linearizzato soddisfi tutte le richieste illustrate nell'*Appendice* A, non esistono rami che biforcano a stato stazionario.

Se guardiamo piú da vicino la differenza tra i modelli considerati nel capitolo III, §§ 5 e 6, e il modello B.2, ci accorgiamo che nel primo, la radice dell'equazione caratteristica, almeno nel punto di stabilità marginale, è semplice, mentre nell'ultimo è sempre doppia. Questa non è una correlazione fortuita, ma piuttosto un esempio del seguente teorema generale:

> Se un autovalore dell'equazione linearizzata attorno a un ramo X_s ha *molteplicità dispari* per $\lambda \simeq \lambda_c$ (cioè se ω è una radice multipla dell'equazione caratteristica con molteplicità dispari) e diventa zero per $\lambda = \lambda_c$, allora esiste almeno un ramo di soluzioni che esce da (X_s, λ_c). Inoltre questo ramo o si estende all'infinito oppure incontra un altro punto di biforcazione.

Si possono ottenere informazioni considerevolmente piú esaurienti nel caso di un autovalore semplice e sotto l'ipotesi che in λ_c non solo si annulli $Re\omega$, ma che il ramo $Re\omega$ come una funzione di λ attraversi l'asse delle λ. Questa *condizione di trasversalità*,

(B.2) $$\frac{d}{d\lambda} Re\omega(\lambda)\Big|_{\lambda = \lambda_c} \neq 0$$

è illustrata nella figura A.I. Si può quindi stabilire il seguente risultato:

Se un ramo (X_s, λ) asintoticamente stabile per $\lambda < \lambda_c$ perde la sua stabilità attraverso un autovalore semplice in $\lambda = \lambda_c$ ed è soddisfatta la condizione di trasversalità in questo valore critico, allora:

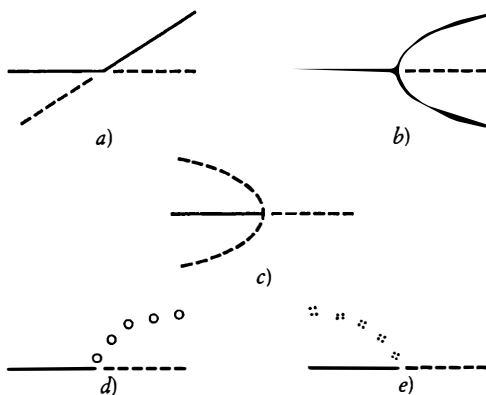
1. i rami che biforcano sopra il valore critico sono stabili, e i rami che biforcano sotto il valore critico sono instabili.
2. le soluzioni che biforcano da uno stato stazionario saranno stazionarie se $Im\omega(\lambda_c) = 0$, e periodiche nel tempo se $Im\omega(\lambda_c) \neq 0$.

La figura B.I riassume le varie possibilità. Non faremo altri commenti su di esse, poiché sono state ampliamente illustrate nel capitolo III, e nell'*Appendice* A.

2. *Sviluppo delle soluzioni in serie di perturbazione.*

Supponiamo ora che siano soddisfatte le condizioni di validità dei precedenti teoremi generali, e descriviamo il procedimento di costru-

Figura B.I

Diagrammi di biforcazione attorno ad uno stato stazionario nel caso di un solo autovalore dell'operatore della stabilità lineare. (*a*): biforcazione transcritica delle soluzioni di stato stazionario (equazione B.12); (*b*) e (*c*): biforcazione simmetrica sopracritica e sottocritica delle soluzioni di stato stazionario (equazione B.13); (*d*): biforcazione di Hopf; (*e*): biforcazione di Hopf sottocritica.

zione della forma esplicita dei rami che biforcano. Per semplificare i particolari tecnici ci limiteremo a considerare biforcazioni di soluzioni di stato stazionario con un autovalore semplice [$Im\,\omega(\lambda_c) = 0$] che nascono come risultato della prima instabilità dello stato di riferimento. Il problema da risolvere è quindi (cfr. equazione B.1):

$$\mathcal{L}(\lambda)\cdot \boldsymbol{x} + \boldsymbol{h}(\boldsymbol{x},\lambda) = 0 \tag{B.3}$$

Abbiamo ripetutamente sottolineato le difficoltà che nascono nel risolvere problemi non-lineari. Per questa ragione abbandoniamo l'idea di ottenere risultati esatti di carattere globale, e limitiamo la nostra attenzione al *comportamento locale* delle soluzioni nelle vicinanze del punto di biforcazione λ_c. Inoltre, supponiamo che le nuove soluzioni nascano in modo continuo in λ_c, escludendo cosí ramificazioni verticali o balzi.

Queste ipotesi ci permettono di sviluppare $\boldsymbol{x}$ in un intorno di λ_c, in serie di potenze di un piccolo parametro. Quest'ultimo deve essere sicuramente legato a $\lambda - \lambda_c$, poiché per $\lambda = \lambda_c$ il valore assoluto della soluzione $|\boldsymbol{x}|$ tende a zero. Non c'è nessuna ragione comunque perché questo piccolo parametro sia lo stesso $\lambda - \lambda_c$, poiché nulla ci garantisce in principio l'analiticità delle soluzioni in $\lambda - \lambda_c$. Effettivamente, nel capitolo III abbiamo incontrato diversi esempi di dipendenza manifestamente non-analitica (cfr. per esempio equazione 3.11*b*).

Per permettere al sistema di fissare esso stesso la dipendenza dalle soluzioni da $\lambda - \lambda_c$, introduciamo un piccolo parametro ausiliario ε rispetto al quale sviluppiamo entrambi $\boldsymbol{x}$ e $\lambda - \lambda_c$:

$$\begin{aligned} \boldsymbol{x} &= \varepsilon \boldsymbol{x}_1 + \varepsilon^2 \boldsymbol{x}_2 + \cdots \\ \lambda - \lambda_c &= \varepsilon\gamma_1 + \varepsilon^2\gamma_2 + \cdots \end{aligned} \tag{B.4}$$

Sostituendo nelle equazioni B.3, otteniamo, al primo ordine in ε,

$$\mathcal{L}(\lambda_c)\cdot \boldsymbol{x}_1 = 0 \tag{B.5}$$

Questo problema omogeneo è stato discusso in dettaglio nell'*Appendice* A in connessione all'analisi della stabilità lineare. La sua soluzione è della forma (si veda l'equazione A.12 con $\omega = 0$, il valore di stabilità marginale per $\lambda = \lambda_c$):

$$\boldsymbol{x}_1 = c\boldsymbol{u} \tag{B.6}$$

Per un uso successivo abbiamo fattorizzato l'ampiezza c della soluzione, la quale a questo livello è indeterminata a causa del fatto che le equazioni B.5 sono omogenee. Si intende pertanto d'ora in poi che $\boldsymbol{u}$ è completamente determinato dal problema linearizzato. Per esempio, per il modello del brussellatore abbiamo da A.8:

$$(B_c - 1)u_1 + A^2 u_2 - D_1 k_c^2 u_1 = 0$$
$$-B_c u_1 - A^2 u_2 - D_2 k_c^2 u_2 = 0$$

dalle quali il vettore $\boldsymbol{u}$ può essere semplicemente calcolato

$$\boldsymbol{u} = \begin{bmatrix} 1 \\ -B_c/(A^2 + D_2 k_c^2) \end{bmatrix}$$

Passiamo ora a ordini piú grandi nell'analisi della perturbazione che ci dovrebbero permettere, in particolare, di fissare l'ampiezza della parte dominante della soluzione, $\varepsilon \boldsymbol{x}_1$. All'ordine ε^2 otteniamo dalle equazioni B.3 tre tipi di contributi. Primo, l'operatore $\mathcal{L}(\lambda_c)$ può agire sul termine del secondo ordine $\boldsymbol{x}_2$ dello sviluppo di $\boldsymbol{x}$. Secondo, possiamo calcolare l'operatore $\mathcal{L}(\lambda)$ in un λ vicino a λ_c, e avere che agisce su $\boldsymbol{x}_1$. Questo spostamento sarà espresso formalmente dal primo termine dello sviluppo di Taylor di $\mathcal{L}$ attorno a λ_c e si leggerà quindi come $\mathcal{L}_\lambda(\lambda_c)\cdot \boldsymbol{x}_1$, dove denotiamo le derivate con dei pedici. Come esempio, se si usa B come parametro di biforcazione nel modello del brussellatore, vediamo dalle equazioni A.8 che

$$\mathcal{L}_\lambda(\lambda_c)\cdot \boldsymbol{x}_1 \big|_{\text{brussellatore}} = \begin{pmatrix} x_1 \\ -x_1 \end{pmatrix}$$

Infine, all'ordine ε^2 avremo un contributo che viene dalla parte quadratica di $\boldsymbol{b}$, che sarà ancora denotata formalmente come il secondo termine di uno sviluppo di Taylor attorno allo zero. Per il modello del brussellatore questo sarà semplicemente la parte quadratica dell'equazione A.6*a*:

$$\begin{bmatrix} (B_c/A)x^2 + 2Axy \\ (-B_c/A)x^2 - 2Axy \end{bmatrix}$$

Denotando ancora le derivate con dei pedici otteniamo infine l'intera equazione all'ordine ε^2:

$$\mathcal{L}(\lambda_c)\cdot \boldsymbol{x}_2 = -\gamma_1 \mathcal{L}_\lambda(\lambda_c)\cdot \boldsymbol{x}_1 - \frac{1}{2}\boldsymbol{b}_{xx}(\lambda_c)\cdot \boldsymbol{x}_1\boldsymbol{x}_1 \tag{B.7}$$

Procedendo esattamente allo stesso modo, otteniamo per gli ordini piú alti:

$$\mathcal{L}(\lambda_c)\cdot x_3 = -\gamma_1\mathcal{L}_\lambda(\lambda_c)\cdot x_2 - \gamma_2\mathcal{L}_\lambda(\lambda_c)\cdot x_1$$

(B.8)
$$-\frac{\gamma_1}{2}h_{xx\lambda}(\lambda_c)\cdot x_1x_1 - \frac{\gamma_1^2}{2}\mathcal{L}_{\lambda\lambda}(\lambda_c)\cdot x_1$$

$$-\frac{1}{6}h_{xxx}(\lambda_c)\cdot x_1x_1x_1 - h_{xx}(\lambda_c)\cdot x_1x_2$$

ecc.

3. *Le equazioni di biforcazione.*

Abbiamo visto che, grazie allo sviluppo in serie della perturbazione, il problema non-lineare iniziale (equazioni B.4) è stato sostituito da una sequenza infinita di problemi *lineari*. Il primo di questi (equazione B.5) è sia lineare che omogeneo, e la sua soluzione a meno di un fattore moltiplicativo è già data dall'equazione B.6. Le rimanenti sono non omogenee, poiché la loro parte destra dipende dalla soluzione delle equazioni degli ordini piú bassi. Ricordando che $\mathcal{L}(\lambda_c)$ ha la struttura di una matrice nello spazio dei vettori $\{x_i\}$, siamo tentati di scrivere la loro soluzione come

(B.9)
$$x_i \sim \mathcal{L}^{-1}(\lambda_c)\cdot q_i \qquad i \geqslant 2$$

dove q_i denotano i membri a destra delle equazioni B.7, B.8, ecc. Il punto però è che $\mathcal{L}(\lambda_c)$ *non* è invertibile, poiché, per l'equazione B.5, esistono soluzioni non banali sulle quali essa vale zero. Un'analoga situazione si incontra in algebra, quando in un insieme di equazioni lineari non omogenee il determinante dei coefficienti delle incognite si annulla. Intuitivamente ci aspettiamo che in questi casi l'azione dell'operatore inverso in accordo con l'equazione B.9 produca risultati diversi, a meno che le parti di q_i che appartengono allo «spazio nullo» di $\mathcal{L}(\lambda_c)$ (lo spazio delle funzioni che soddisfano l'equazione B.9) non possano essere eliminati. Un importante risultato di analisi, noto come il teorema dell'*alternativa di Fredholm*, ci dice come questo si può ottenere. Brevemente, esso afferma che il membro a destra q_i debba essere *ortogonale* ai vettori u^+ che soddisfano l'equazione

(B.10)
$$\mathcal{L}^+(\lambda_c)\cdot u^+ = 0$$

nella quale $\mathcal{L}^+$ è l'operatore aggiunto di $\mathcal{L}$. Per sistemi di reazione-diffusione $\mathcal{L}^+$ è semplicemente la matrice trasposta di $\mathcal{L}$, poiché il laplaciano ∇^2 è autoaggiunto per condizioni al bordo simmetriche. Ad esempio per il modello del brussellatore:

$$\mathcal{L}^+(\lambda_c) = \begin{pmatrix} B_c - 1 + D_1\nabla^2 & -B_c \\ A^2 & -A^2 + D_2\nabla^2 \end{pmatrix}$$

Come si dovrebbe intendere l'ortogonalità nel precedente teorema? Abbiamo già insistito sul carattere duale del vettore di stato $\boldsymbol{x}$, sia come un vettore dello spazio ordinario, finito dimensionale, sia come una funzione di uno spazio infinito dimensionale. Nello spazio ordinario finito dimensionale l'ortogonalità è espressa dall'annullarsi del prodotto scalare di due vettori, che noi denotiamo con un punto: $\boldsymbol{u}^+ \cdot \boldsymbol{q}_i$. Il prodotto scalare in uno spazio funzionale è qualcosa di piú sottile e qui non sarà spiegato nei dettagli. È sufficiente dire che in molti problemi che sorgono in meccanica quantistica e in altre aree della fisica, si riduce a un integrale sullo spazio del prodotto di funzioni (o di un prodotto scalare di vettori). Denotando con le parentesi $\langle\rangle$ questo prodotto generalizzato, possiamo infine scrivere l'alternativa di Fredholm nella forma

(B.11) $$\langle \boldsymbol{u}^+, \boldsymbol{q}_i \rangle = 0 \qquad i \geqslant 2$$

Il punto è che, proprio come $\boldsymbol{u}$, il vettore $\boldsymbol{u}^+$ è determinato completamente. Inoltre, a parte la sua dipendenza dalle coordinate spaziali, $\boldsymbol{q}_i$ dipende solo dall'ampiezza indeterminata c. L'equazione A.11, nella quale si intende che si deve eseguire un integrale sullo spazio, si riduce quindi a una equazione algebrica per l'ampiezza c del contributo dominante alla soluzione che biforca, $\boldsymbol{x}_1$. Introducendo l'intera ampiezza $z = \varepsilon c$, di $\varepsilon \boldsymbol{x}_1$, con della semplice algebra si ricavano dalla B.11 le seguenti *equazioni di biforcazione*:

(B.12) $$-(\lambda - \lambda_c)Q_1 z + Q_2 z^2 = 0 \qquad \text{per } Q_2 \neq 0$$

(B.13) $$-(\lambda - \lambda_c)Q_1 z + Q_3 z^3 = 0 \qquad \text{per } Q_2 = 0,\ Q_3 \neq 0$$

dove i coefficienti Q_i seguono direttamente dalla nostra analisi. I loro valori espliciti dipendono dalla struttura specifica del sistema considerato.

Si ricorda che siamo partiti da un problema coinvolgente un gran

numero di variabili le quali possono dipendere, in aggiunta, in modo complesso dalle coordinate spaziali. Ciò nonostante nelle vicinanze del punto di biforcazione è stato possibile ridurre la dinamica a delle *forme normali* molto semplici, simili a quelle discusse nel capitolo III, § 5, che coinvolgono solo una quantità incognita, il *parametro d'ordine* z. Questo convalida gli argomenti sviluppati nella seconda parte del capitolo III, § 6.

Equazioni di biforcazione piú intricate sorgono quando si considerano soluzioni periodiche nel tempo o biforcazioni vicino ad autovalori multipli. Per un breve resoconto su questo argomento rimandiamo il lettore all'opera di J. Guckenheimer e Ph. Holmes, *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields* (Berlin 1983).

C

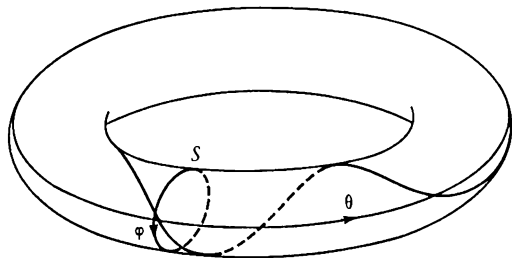
Perturbazione di moti risonanti in sistemi conservativi non-integrabili

1. *La mappa twist.*

Nel capitolo III, § 3, abbiamo visto che il moto di un sistema conservativo integrabile con N gradi di libertà giace in una superficie toroidale N-dimensionale inserita in uno spazio delle fasi a $2N$ dimensioni (cfr. figura 3.7*b* per $N = 2$). Vogliamo capire intuitivamente l'effetto delle perturbazioni non-integrabili su questo tipo di moto e, in particolare, i cambiamenti qualitativi indotti vicino alla risonanza che sono stati anticipati nel capitolo III, § 8. Per permettere una visualizzazione ci limitiamo in questa appendice al caso a due gradi di libertà, ma vogliamo sottolineare che le principali argomentazioni valgono anche per sistemi piú generali.

Siano θ, φ i due angoli che determinano rispettivamente la «longitudine» e la «latitudine» sul toro (figura C.1), ω_1 e ω_2 le frequenze angolari del moto lungo queste due direzioni. Nelle variabili angolo-azione le equazioni del moto sono semplicemente:

$$
\begin{aligned}
I &= I_0 = \text{cost} \\
\varphi &= \omega_1(I)t + \varphi_0 \\
\theta &= \omega_2(I)t + \theta_0
\end{aligned}
\tag{C.1}
$$

Figura C.1
Superficie di sezione di Poincaré S adatta a studiare moti risonanti su di un toro.

Abbiamo mostrato solo una variabile d'azione I, poiché, in virtú della conservazione dell'energia, la seconda viene interamente espressa in termini di I e della energia totale. Inoltre, abbiamo tenuto conto della dipendenza delle frequenze angolari dalla variabile di azione.

È conveniente visualizzare il moto su una superficie di sezione di Poincaré S (si veda anche la figura 3.22) lungo un piano meridionale. Le traiettorie rappresentative intersecheranno questa superficie ogni qualvolta esse eseguiranno un numero intero di giri lungo la direzione θ, per la quale $\theta - \theta_0 = 2\pi n$ $(n = 1, 2, \ldots)$ Le intersezioni si verificheranno dunque ai tempi $t_n = 2\pi n/\omega_2$. Sostituendo questo valore di t nell'equazione per φ otteniamo dunque dalle equazioni C.1:

$$\varphi_n - \varphi_0 = 2\pi \frac{\omega_1}{\omega_2} n$$

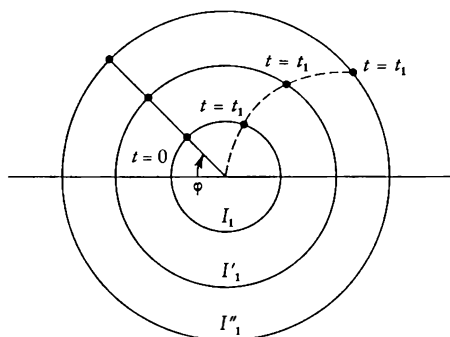
ovvero, introducendo la notazione appropriata per una relazione di iterazione:

$$I_{n+1} = I_n$$
$$\varphi_{n+1} = \varphi_n + 2\pi\alpha(I_n) \qquad \text{(C.2)}$$

dove il numero di rotazione α è definito da

$$\alpha = \frac{\omega_1}{\omega_2} \qquad \text{(C.3)}$$

Figura C.2

La trasformazione twist. Le immagini di tre condizioni iniziali che giacciono sullo stesso raggio per $t = 0$ subiscono una traslazione nella fase per $t = t_1$.

La trasformazione data dalle equazioni C.2 trasforma ovviamente cerchi in cerchi. Tuttavia, il numero di rotazione cambierà in generale da un cerchio a un altro poiché dipende dalla variabile di azione, il cui valore «segnala» i differenti cerchi. In altre parole, partendo da un dato valore φ_0 di φ a $t = 0$, i punti rappresentativi sui diversi cerchi saranno traslati nella fase l'uno rispetto all'altro a ogni tempo futuro $t = t_1$ (si veda la figura C.2). Per questa ragione la trasformazione C.3 viene chiamata *twist mapping*. Come sottolineato nel capitolo III, § 8 in connessione con la figura 3.22, ogni traiettoria periodica sul toro ($\alpha = r/s$ razionale, r e s essendo degli interi) taglia la superficie di Poincaré in punti fissi, il cui numero è uguale al numero s lungo la direzione θ.

Consideriamo ora una debole perturbazione del sottostante sistema hamiltoniano. Questa sarà riflessa da una trasformazione twist perturbata della forma:

$$
\begin{aligned}
I_{n+1} &= I_n + \varepsilon f(I_n, \varphi_n) \\
\varphi_{n+1} &= \varphi_n + 2\pi\alpha(I_n) + \varepsilon g(I_n, \varphi_n) \qquad (\varepsilon \ll 1)
\end{aligned}
\tag{C.4}
$$

Si può mostrare che in conseguenza del teorema di Liouville (si veda il capitolo III, § 2) questa trasformazione conserverà l'area nello spazio delle fasi.

2. *Effetto della perturbazione nel caso di numeri di rotazione razionali.*

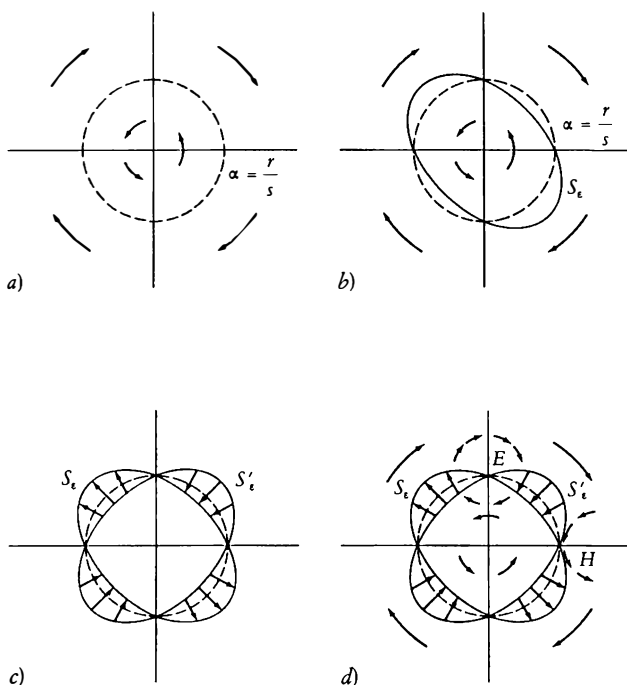
Per ogni trasformazione twist non perturbata (equazioni C.2), la discussione precedente mostra che *ogni* punto sul cerchio I = cost. è un punto fisso di periodo s poiché è obbligato a essere l'intersezione della superficie di Poincaré con una certa traiettoria periodica dopo che quest'ultima ha eseguito s giri lungo la direzione longitudinale. Il risultato principale che riguarda l'effetto della perturbazione in questo quadro è dovuto a Poincaré e Birkhoff. Esso afferma che fuori da questo continuum di punti fissi rimangono $2ks$ punti fissi dopo la perturbazione, con k intero, metà dei quali sono iperbolici e l'altra metà ellittici. Schematizziamo la dimostrazione riferendoci costantemente alle figure C.3*a*-*d*.

Supponiamo per definitezza che la direzione lungo cui α aumenta sia quella diretta verso l'interno. Allora il sistema integrabile possiederà necessariamente due traiettorie quasi-periodiche, una esterna e una interna al toro risonante che dopo s giri verranno trasformate sulla super-

ficie di Poincaré rispettivamente in senso orario e antiorario, a seconda delle posizioni del loro passaggio iniziale (cfr. figura C.3*a*).

Poiché il moto quasi-periodico rimane robusto per piccole perturbazioni, vi saranno ancora due curve del sistema perturbato che possiedono le proprietà di cui sopra dopo s iterazioni della trasformazione twist perturbata. Fra loro vi sarà dunque una curva che rappresenta l'intersezione di una famiglia di traiettorie del sistema perturbato con la super-

Figura C.3

Esempio del teorema di Poincaré-Birkhoff per il quale alcuni punti fissi sulla superficie di Poincaré S si preservano quando un moto risonante viene leggermente perturbato. *a*): le frecce ai due lati della curva tratteggiata indicano la direzione delle traslazioni di fase subite da due traiettorie quasi periodiche attraverso successive iterazioni della trasformazione; *b*): luogo di intersezione delle traiettorie del sistema perturbato che ritornano alla stessa coordinata angolare φ dopo s iterazioni con la superficie di Poincaré (curva S_ε), dove s è il numero di rotazione lungo l'angolo θ; *c*): immagine di S_ε (curva $S_{\varepsilon'}$) dopo s iterazioni della trasformazione perturbata. Si noti la presenza di un numero pari di punti fissi; *d*): stabilità dei punti fissi della mappa twist perturbata. *E*: punti ellittici, *H*: punti iperbolici.

ficie di Poincaré (indicata con S_ε nella figura C.3*b*) la cui coordinata angolare φ rimane invariata dopo s iterazioni della trasformazione twist perturbata. Tuttavia, rimangono possibili i cambiamenti nella direzione *radiale*. Tipicamente, dunque, dopo s iterazioni la curva S_ε verrà spostata. Indichiamo con S_ε' l'immagine di S_ε generata in questo modo (si veda la figura C.3*c*).

Come sottolineato precedentemente, a causa del suo carattere conservativo la trasformazione conserva l'area nella superficie di Poincaré. Ne segue che le curve S_ε' e S_ε devono racchiudere un'area uguale. Ma ciò è possibile solo se esse si intersecano l'un l'altra un numero pari di volte, diciamo $2k$. Per costruzione, ogni punto dell'incrocio ritorna alla sua posizione iniziale quando viene iterato s volte. È dunque un punto fisso di periodo s. Ciascuna delle sue s iterazioni, che insieme costituirebbero la soluzione di periodo s del sistema imperturbato, sarà naturalmente anche un punto fisso. Abbiamo stabilito in questo modo l'esistenza di punti fissi il cui numero è uguale a un multiplo pari del numero di avvolgimento s. Il ragionamento seguito non dà informazioni sulla grandezza dell'intero k.

Procediamo ora alla discussione della stabilità di questi punti fissi. Consideriamo le vicinanze del punto E della figura C.3*d*. Una perturbazione fuori dal punto fisso, diciamo nella direzione verso il basso, porterà dapprima a una rotazione in senso orario che nasce dal cambiamento del numero di rotazione α, quindi a un moto verso l'interno sotto l'effetto della trasformazione (seguire le frecce nella figura), e poi a una rotazione in senso antiorario, e infine a un moto verso l'esterno. Il risultato globale è una rotazione attorno a E, che si comporta dunque come un punto fisso ellittico (stabile). D'altra parte, una perturbazione lontana dal punto H mostrato nella figura C.3*d* porta il punto rappresentativo sempre piú lontano da H, che si comporta dunque come un punto fisso iperbolico (instabile). Questa alternanza di punti fissi di diversa stabilità fornisce una spiegazione qualitativa degli aspetti principali del diagramma di fase mostrato nella figura 3.23.

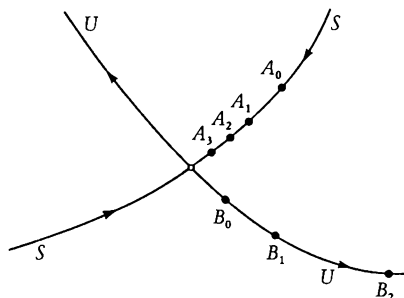
3. *Punti omoclinici.*

Cosa succede quando compare un punto fisso instabile di tipo iperbolico (un punto di sella) nella mappa di Poincaré? Abbiamo visto nel capitolo III, §§ 3-5, che, nelle vicinanze di un tale punto, esistono due traiettorie eccezionali, le separatrici, che si uniscono proprio a questi

punti (figura C.4, curve S e U). Ciò rimane compatibile con l'unicità delle soluzioni del sottostante sistema dinamico (cfr. capitolo III, § 1) finché qualche condizione iniziale su S o U non può raggiungere il punto fisso H in un tempo finito. Cosí, se consideriamo una condizione iniziale A_0 sulla separatrice convergente S, la sua prima iterazione con la trasformazione porterà a uno stato A su S piú vicino a H, ma la sequenza $A_0, A_1, \ldots, A_n$ convergerà a H solo per $n \to -\infty$. Analogamente, una condizione iniziale B_0 sulla separatrice divergente U darà luogo per tempi positivi a un insieme di stati B_1, B_2. E sempre piú lontani da H, mentre per tempi negativi darà luogo a una sequenza simile a A_n, che converge a H, quando $n \to \infty$. La questione che solleviamo ora è ciò che accade alle separatrici lontane dal punto fisso H, specialmente quando, come vedemmo piú sopra in questa appendice, H coesiste con molti altri punti fissi iperbolici.

La figura C.5 rappresenta tre possibilità: nel primo caso le separatrici uniscono due dei punti iperbolici coesistenti in un tale modo che la separatrice stabile di un punto diventa la separatrice instabile dell'altro, e viceversa. Tali *traiettorie eterocliniche* appaiono già nei sistemi Hamiltoniani integrabili come il pendolo (si veda la figura 3.8). Nel secondo caso le due curve S e U si identificano, e ciò risulta in una *traiettoria omoclinica*. Abbiamo già alluso a questa possibilità nel capitolo III, § 10 (cfr. figura 3.29). Una terza possibilità, illustrata nella figura C.5*c*, è semplicemente che la separatrice instabile tende all'infinito e quella stabile viene dall'infinito.

Ora, il lettore comprenderà che le prime due situazioni sono piuttosto eccezionali, mentre la terza è fisicamente non realistica poiché im-

Figura C.4
Successive iterazioni dei punti A_0 e B_0 che giacciono rispettivamente sulla separatrice stabile ed instabile di un punto fisso iperbolico.

plica (in assenza di altri punti fissi attrattivi) il comportamento esplosivo di una certa variabile che descrive lo stato del sistema. La figura C.6 descrive una situazione molto piú naturale, scoperta da Poincaré: le separatrici stabili e instabili che partono dal punto fisso H si intersecano in un punto P, per il quale Poincaré coniò il nome di *punto omoclinico*. Enfatizziamo ancora che una tale intersezione non è in contraddizione con l'unicità delle soluzioni del sottostante sistema dinamico continuo, anche se P non è un punto fisso della mappa, poiché U e S *non* rappresentano le traiettorie di fase del sistema.

Ora, la figura C.6 nasconde un ricco comportamento di incredibile complessità. Poiché P appartiene alla separatrice stabile S, le sue iterazioni $P_1, P_2, \dots, P_n$ per tempi postivi giaceranno tutti su S e convergeranno al punto fisso H per $n \to \infty$. D'altra parte, P è anche contenuto in una separatrice instabile U, ed è dunque l'immagine di certi punti della separatrice quando il sistema evolve sulla mappa di Poincaré per successive iterazioni. Alternativamente, questi punti si possono pensare come le iterazioni $P_{-1}, P_{-2}, \dots, P_{-n}$ di P per tempi negativi, che convergono al punto fisso H per $n \to \infty$. Ma poiché P viene piú tardi rigettato su P_1,

Figura C.5

Configurazioni possibili delle separatrici dei punti fissi iperbolici. *a*): Orbite eterocliniche; *b*): orbita omoclinica (cappio della separatrice); *c*): separatrici che si estendono all'infinito.

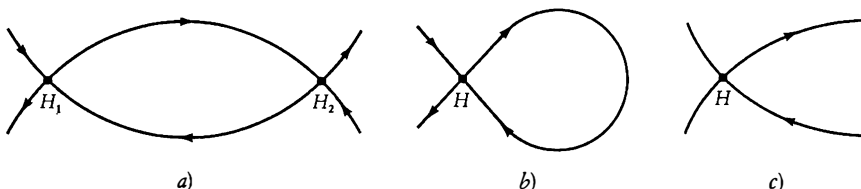
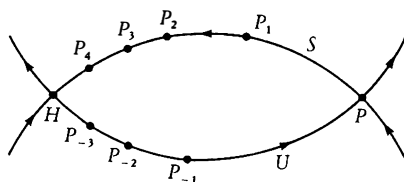
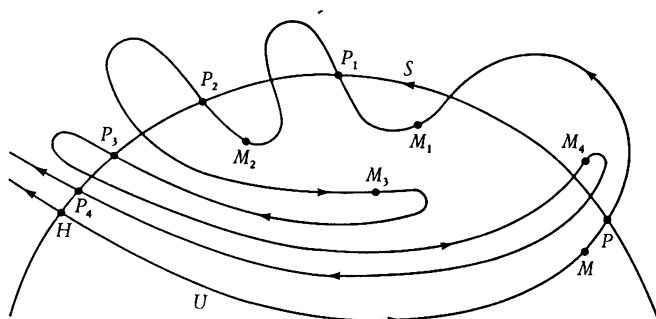



Figura C.6

Punto omoclinico P, in una superficie di sezione di Poincaré: separatrici stabili ed instabili si intersecano trasversalmente in P., P_1, P_2, ... rappresentano iterazioni successive di P per tempi positivi e giacciono su una varietà stabile S, mentre le preimmagini P_{-1}, P_{-2}, ... di P giacciono su una varietà instabile U.

P_2, ..., P_n tutti questi punti verranno trasformati prima o poi anche su P_1, P_2, ..., P_n. Raggiungiamo quindi la notevole conclusione che la curva U attraversa i punti P_1, P_2. È, P_n della curva S. In altre parole, P_1, P_2, ..., P_n sono anche punti omoclinici e tali sono anche, naturalmente, P_{-1}, P_{-2}, ..., P_{-n}, ... da un singolo punto omoclinico P abbiamo ottenuto una doppia infinità di tali punti!

Rimane da vedere come le curve U e S si adatteranno a questa situazione. Chiaramente esse dovranno piegarsi e ondeggeranno continuamente, con una frequenza crescente vicino ai punti fissi iperbolici. Ci si può convincere che queste ondulazioni sono associate a escursioni sempre piú larghe dal punto fisso, seguite da reinserimenti di ritorno alle sue vicinanze, come indicato nella figura C.7. Questo spiega qualitativamente l'affermazione fatta alla fine del capitolo III, § 9, ovvero, che la presenza di punti omoclinici dà luogo a dinamiche simili alla mappa a ferro di cavallo. Questo suggerisce a sua volta la possibilità di moti instabili di tipo caotico, in accordo con i risultati delle soluzioni numeriche delle intere equazioni di evoluzione non-lineari dei sistemi dinamici conservativi non-integrabili o dei sistemi dinamici dissipativi.

Figura C.7

Effetto di un punto omoclinico sulla genesi del moto complesso vicino ad una separatrice. Varietà stabili ed instabili del punto iperbolico si intersecano infinite volte. Nei sistemi conservativi, la conservazione delle aree implica inoltre che queste separatrici ondeggiano fortemente nella vicinanza di H. Dopo un numero sufficiente di iterazioni, un punto M inizialmente ad un lato della separatrice stabile sarà respinto nell'altro lato (punto M_4 nella figura). È quindi soddisfatto uno dei prerequisiti per l'inizio della casualità.

D

Ricostruzione della dinamica dei sistemi complessi da dati in serie temporale: applicazione alla variabilità climatica

1. *Considerazioni introduttive.*

Sperimentalmente, per indagare il comportamento di un sistema complesso si osserva, durante un certo periodo di tempo, una variabile che si considera essere «pertinente». Un esempio di una tale *serie temporale* è fornito dalla figura 1.17, nella quale è rappresentato il grafico in funzione del tempo, dei cambiamenti del volume totale di ghiaccio, dedotti dai dati dell'isotopo $^{18}O/^{16}O$ dei sedimenti del nucleo delle profondità oceaniche. Un altro esempio è l'elettroencefalogramma (EEG), la somma delle fondamentali attività autosostenute dei neuroni di frequenza relativamente bassa (dell'ordine di 0.5 fino a 40 Hz) che scaturiscono da piccoli volumi del tessuto corticale appena sotto il cuoio capelluto.

A una prima occhiata, una serie temporale di una variabile individuale sembra fornire una quantità piuttosto limitata di informazioni. In particolare, si può arguire che essa è ristretta a una visione «monodimensionale» del sistema, il quale, in realtà, contiene un grande numero di variabili indipendenti tra loro. In questa *Appendice* mostriamo che una serie temporale contiene una quantità di informazioni molto piú ricca: essa serba le tracce di tutte le altre variabili partecipanti alla dinamica, e permette di identificare alcuni aspetti fondamentali del sottostante sistema indipendentemente da ogni modellistica.

Sia $X_0(t)$ l'insieme dei valori delle variabili disponibili dalla serie temporale. Naturalmente, di fatto ci si aspetta che altre variabili $\{X_k(t)\}$, $k = 1, \ldots, n-1$, intervengano nella dinamica. Comunque, poiché non ci basiamo su alcun modello particolare, vogliamo ricostruire questa dinamica solamente dalla conoscenza di $X_0(t)$. Per questo fine consideriamo lo spazio delle fasi generato dalle variabili $\{X_k\}$, $k = 0, \ldots, n-1$. Come abbiamo visto nel capitolo III uno stato istantaneo del sistema è rappresentato in questo spazio da un punto, chiamiamolo P, mentre

una sequenza nel tempo di questi stati definisce una curva, la traiettoria dello spazio delle fasi. Con il crescere del tempo il sistema raggiunge uno stato di regime permanente, se la sua dinamica può essere ricondotta a un insieme di leggi deterministiche. Questo si riflette sulla convergenza verso un sottoinsieme dello spazio delle fasi di famiglie di traiettorie di fase. Abbiamo ripetutamente chiamato questo sottoinsieme invariante l'*attrattore.* Nostro scopo principale qui è quello di cercare una risposta alle seguenti domande:

1. È possibile identificare un attrattore per una data serie temporale? In altre parole, si possono considerare gli aspetti salienti del sistema osservato attraverso questa serie temporale come la manifestazione di una dinamica deterministica (possibilmente molto complessa) o piuttosto essi contengono un elemento irriducibilmente stocastico?

2. Supponendo che esista un attrattore, qual è la sua dimensione d? Sappiamo dall'analisi contenuta nel capitolo III che quest'ultima ci fornirebbe una considerevole quantità di informazione sulla dinamica del sistema. Per esempio, se $d = 1$ stiamo trattando oscillazioni autosostenute *periodiche*; se $d = 2$ siamo in presenza di oscillazioni *quasi periodiche* con due frequenze incommensurabili; e se d è maggiore di due e non è un intero (un attrattore frattale, si veda il capitolo III, § 10), il sistema esibisce una oscillazione caotica che evidenzia una grande sensibilità nei confronti delle condizioni iniziali, cosí come una imprevedibilità intrinseca.

3. Qual è la dimensione minima n dello spazio delle fasi nel quale si può immergere il sopra citato attrattore? Questa definisce il numero minimo di variabili che si devono considerare nella descrizione della dinamica sottostante. Si noti che d è necessariamente piú piccolo di n.

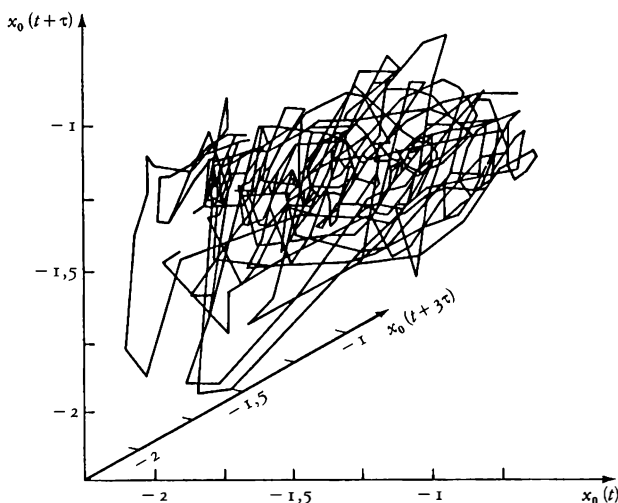
Il nostro primo passo consisterà nell'identificare un opportuno insieme di variabili che genera questo spazio delle fasi. Un modo molto conveniente è quello di estendere l'originale serie temporale $X_0(t)$ attraverso successive traslazioni sempre piú grandi definite come multipli interi di un fissato sfasamento τ ($\tau = m\Delta t$, dove m è un intero e Δt è l'intervallo tra le successive campionature). Prendendo inoltre N punti

equidistanti dall'insieme dei dati, siamo condotti al seguente insieme di variabili discretizzate:

$$
\begin{aligned}
&X_0: \quad X_0(t_1), \ldots, X_0(t_N)\\
&X_1: \quad X_0(t_1+\tau), \ldots, X_0(t_N+\tau)\\
&\vdots\\
&X_{n-1}: \quad X_0[t_1+(n-1)\tau], \ldots, X_0[t_N+(n-1)\tau]
\end{aligned}
\qquad (D.1)
$$

Per una caratteristica scelta di τ ci si aspetta che queste variabili siano linearmente indipendenti, e ciò è tutto quello che occorre per definire propriamente uno spazio delle fasi. Inoltre tutte queste variabili possono essere dedotte dalla serie temporale individuale relativa a $X_0(t)$, come fornito dai dati. Vediamo pertanto che, in principio, abbiamo a disposizione una sufficiente quantità di informazione per andare oltre lo spazio monodimensionale della serie temporale originaria e per esten-

Figura D.1

Ritratto delle fasi che rappresenta l'evoluzione climatica durante l'ultimo milione di anni compreso in uno spazio tridimensionale. Il risultato utilizza circa 500 valori equidistanti di X_0, la variabile descritta nella figura 1.17. Essi sono ricavati dalla documentazione degli isotopi di ossigeno ottenuti dal nucleo della profondità abissale V 28-238 di Shackleton e Opdyke, utilizzando una interpolazione disponibile dalla banca dei dati dell'Università di Lovanio. Il valore della traslazione τ usata nel disegnare la figura è di τ = 2000 anni.

dere la dinamica del sistema in uno spazio delle fasi multidimensionale. Questa quantità di informazioni ci permette già di disegnare il *ritratto delle fasi* del sistema o, piú precisamente, la sua proiezione in un sottospazio di dimensione piú bassa dell'intero spazio delle fasi. La figura D.1 descrive un ritratto delle fasi tridimensionale dell'evoluzione climatica durante l'ultimo milione di anni dedotto dai dati della figura 1.17. Questa prospettiva mostra chiaramente la complessità del moto sottostante. Comunque i dati sono troppo grossolani per trarre alcuna conclusione solamente da questa figura. Il nostro prossimo compito è quello di caratterizzare piú profondamente la complessità della dinamica, usando le tecniche della teoria dei sistemi dinamici.

2. *Retroterra teorico per l'analisi dei dati.*

Introduciamo la notazione vettoriale: X_i denota un punto dello spazio fasi le cui coordinate sono $\{X_0(t_i), \ldots, X_0[t_i + (n-1)\tau]\}$. Si sceglie ora tra questi dati un punto di «riferimento» X_i e si calcolano tutte le distanze $|X_i - X_j|$ dagli $N - 1$ punti rimanenti. Questo ci permette di calcolare i punti dei dati che sono all'interno di una prescritta distanza, r dal punto X_i dello spazio delle fasi. Ripetendo il processo per tutti i valori di i si arriva alla quantità

$$C(r) = \frac{1}{N^2} \sum_{\substack{i,j=1 \\ i \neq j}}^{N} \theta(r - |X_i - X_j|) \qquad \text{(D.2)}$$

dove θ è la funzione di Heaviside, $\theta(x) = 0$ se $x<0$ e $\theta(x) = 1$ se $x>0$. Il non annullarsi di $C(r)$ misura il grado in cui la presenza di un punto dei dati X_i influisce sulla posizione degli altri punti. $C(r)$ può dunque essere pensata come la *funzione di correlazione* intera dell'attrattore.

Supponiamo di fissare un ε piccolo e di usarlo come un parametro per sondare la struttura dell'attrattore. Se quest'ultimo è una linea, chiaramente il numero dei punti dei dati entro una distanza r da un punto prescritto dovrebbe essere proporzionale a r/ε. Se esso è una superficie, questo numero dovrebbe essere proporzionale a $(r/\varepsilon)^2$ e, piú generalmente (cfr. capitolo III, § 7), se esso è una varietà a d dimensioni sarà proporzionale a $(r/\varepsilon)^d$. Ci aspettiamo dunque che per r relativamente piccolo $C(r)$ varierà come

$$C(r) \sim r^d \qquad \text{(D.3)}$$

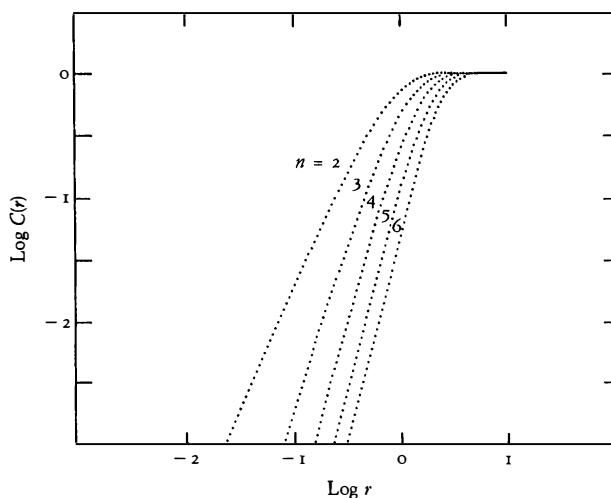
In altre parole, la dimensionalità d dell'attrattore è data dalla pendenza di $\ln C(r)$ in funzione di $\ln r$ entro un certo dominio di valori di r:

$$\ln C(r) = d \ln r \tag{D.4}$$

I risultati di cui sopra suggeriscono il seguente algoritmo:

1. Partendo da una serie temporale fornita dal EEG, costruire la funzione di correlazione (equazione D.2) considerando successivamente valori piú alti della dimensione n dello spazio delle fasi.
2. Dedurre la pendenza d vicino all'origine secondo l'equazione D.4 e vedere come cambia il risultato all'aumentare di n.
3. Se la dipendenza di d da n si satura oltre un qualche n relativamente piccolo, il sistema rappresentato dalla serie temporale deve possedere un attrattore. Il valore di saturazione d è visto come la dimensionalità dell'attrattore rappresentato dalla serie temporale. I valori di n oltre il quale si osserva la saturazione fornisce il numero minimo di variabili necessarie per modellare il comportamento rappresentato dall'attrattore.

Figura D.2

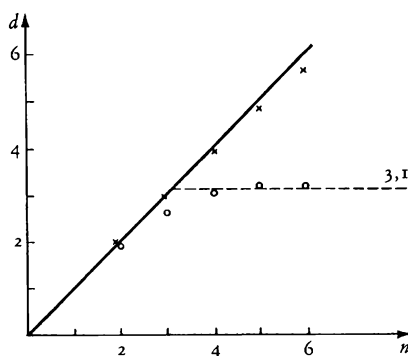
Dipendenza dalla distanza della funzione correlazione sull'attrattore climatico. Si noti l'esistenza di un esteso dominio di linearità, dal quale si può dedurre la dimensione dell'attrattore. In questo caso τ = 8000 anni.

3. *L'attrattore climatico.*

Questa procedura è stata applicata per costituire l'analisi dei dati climatici pertinenti alla figura 1.17. La figura D.2 mostra la dipendenza di ln $C(r)$ in funzione di ln r per $n = 2$ fino a $n = 6$. Vediamo che c'è in effetti una regione estesa lungo la quale questa presenza è lineare, in accordo con l'equazione D.4.

La figura D.3 mostra (punti circolari) che la pendenza raggiunge un valore di saturazione per $n = 4$, che è circa $d \simeq 3.1$.

Lo stesso disegno mostra come varia d rispetto a n se il segnale considerato è rumore bianco gaussiano: non c'è la tendenza alla saturazione. Infatti in questo caso d risulta essere uguale a n. Questo stabilisce una differenza netta tra la dinamica climatica e il rumore casuale e mostra come l'evoluzione climatica dell'ultimo milione di anni possa essere vista come la manifestazione di una dinamica deterministica che possiede un attrattore di bassa dimensione. Il fatto che la dimensione dell'attrattore sia frattale ci fornisce un modo naturale per capire la variabilità intrinseca e l'imprevedibilità del sistema climatico, poiché entrambi questi aspetti si annoverano fra le principali caratteristiche della dinamica caotica. Un fatto notevole è come questa dinamica sia riconducibi-

Figura D.3

Dipendenza della dimensione d dal numero delle variabili dello spazio delle fasi n per l'attrattore climatico (∘) e per un segnale di rumore bianco (×) per $\tau = 8000$ anni. Si noti la saturazione a un valore di plateau $\sim 3,1$ nel primo caso e la relazione $d \sim n$ nel secondo caso.

le a un insieme limitato (esattamente 4) di variabili chiave, sebbene in questo stadio non possa essere identificata la loro specifica natura.

È istruttivo confrontare questa conclusione che, in qualche modo, ci è stata imposta dai dati, e i risultati estratti dai modelli tratteggiati brevemente nel capitolo VI, § 4. Emergono due punti principali. Primo, l'instabilità del moto sull'attrattore caotico permesso dall'amplificazione di deboli segnali esterni, come le piccole variazioni dell'energia solare che nascono dalla variabilità dell'orbita della terra. La possibilità di tale amplificazione fu infatti una delle conclusioni ricavate nel capitolo VI, § 4, sebbene il fenomeno venne attribuito agli effetti stocastici che inducevano transizioni continue tra gli stati coesistenti.

Il secondo punto concerne l'origine delle forzature stocastiche che si devono aggiungere all'equazione del bilancio dell'energia (si veda l'equazione 6.7. Come visto nel capitolo IV, §§ 7 e 8, un attrattore caotico può generare una varietà di processi stocastici. Possiamo considerare dunque le fluttuazioni nella dinamica del clima come una delle manifestazioni del carattere caotico dell'attrattore sottostante.

4. *Conclusioni e prospettive.*

Il precedente approccio è stato anche applicato ai dati del EEG che corrispondono a vari stati di sonno. Di nuovo l'esistenza di un attrattore per questo sistema sembra essere assicurata, almeno nello stadio di sonno profondo. Dipendentemente dal soggetto, la sua dimensione risulta essere circa tra 4 e 4,3. Alla luce dei commenti fatti nel capitolo IV, §§ 7 e 8, si è tentati di collegare questo risultato alla capacità del cervello di generare ed elaborare l'informazione. Sarebbe importante analizzare dallo stesso punto di vista altri sistemi complessi che esistono in natura, come per esempio le attività economiche.

Un numero di altre proprietà di grande importanza qualitativa e quantitativa, come le funzioni di correlazione e gli esponenti di Ljapunov possono anche essere determinati solamente dai dati delle serie temporali. Non discutiamo ora questi punti nel dettaglio ma rimandiamo il lettore alle pubblicazioni originali.

E

Processi primordiali

1. *Introduzione.*

Come si è sottolineato nel capitolo 1 una teoria della autoorganizzazione non può essere completa se essa non prende in considerazione la gravitazione, la piú universale fra tutte le forze. La teoria della relatività generale forma la base della moderna teoria della gravitazione. È stata definita da Landau e Lifschitz «la piú bella di tutte le teorie esistenti», e corrisponde certamente a una drastica revisione della descrizione newtoniana del nostro universo. La visione newtoniana era basata su una dualità: da una parte, lo spazio-tempo; dall'altra, la materia (lo spazio e il tempo erano considerati indipendenti uno dall'altro cosí come dal contenuto di materia). La novità fondamentale della relatività generale di Einstein fu di sopprimere questa dualità e di stabilire una connessione fra lo spazio-tempo da una parte e la materia dall'altra.

Le equazioni di campo di Einstein mettono in relazione le proprietà dello spazio-tempo con le proprietà della materia; le prime sono essenzialmente correlate alla curvatura, mentre le seconde sono correlate alla pressione e la densità di energia. In quest'ottica sono possibili molti spazi-tempo: lo spazio-tempo piatto o gli spazi curvi. Lo spazio-tempo diventa esso stesso un fenomeno fisico, sulla cui origine ed evoluzione possiamo indagare. Questo è infatti uno degli scopi fondamentali della cosmologia moderna. È sorprendente che Einstein, già nel 1917, un anno dopo la formulazione della teoria della relatività generale, presentasse il primo modello teorico di cosmologia. Einstein voleva arrivare a una descrizione dell'universo indipendente dal tempo: per lui l'irreversibilità e la freccia del tempo erano espressioni di soggettività che dovevano essere eliminate dalla scienza. In linea con queste idee, Einstein presentò un modello statico dell'universo.

La successiva evoluzione della cosmologia è uno degli esempi piú sensazionali dell'irruzione del tempo nella fisica moderna. Nel 1923, Alexander Friedmann mostrò che l'universo di Einstein era instabile e propose una teoria di un universo in espansione, piú tardi confermata

dalle osservazioni di Edwin Hubble (1929): le galassie, viste dalla terra, si allontanano con una velocità proporzionale alla loro distanza. Da qui nacque il concetto di universo in evoluzione; si poteva ancora pensare tuttavia che questo fosse un aspetto puramente geometrico, di nessuna importanza per la fisica.

Nel 1965 Penzias e Wilson scoprirono l'esistenza di una radiazione residua di corpo nero: il nostro universo è pieno di fotoni che sono in un equilibrio termico corrispondente alla temperatura di ~2,7K (fu difficile identificare sperimentalmente questa radiazione proprio a causa di una cosí bassa temperatura). La radiazione era stata prodotta durante i primi stadi dell'universo (da cui il termine «residua»): essa appariva come il risultato dell'espansione cosmologica, quando la temperatura era divenuta sufficientemente bassa da permettere la formazione degli atomi: a questo punto la radiazione divenne disaccoppiata dalla materia. Dal momento della scoperta della radiazione residua di corpo nero, sappiamo che l'universo ha una struttura duale: da una parte, le particelle con massa (barioni); dall'altra, particelle senza massa (per esempio, i fotoni). I barioni sono largamente superati in numero dai fotoni: ci sono centinaia di milioni di fotoni per ogni barione. Inoltre è sorprendente che l'entropia collegata con questa radiazione residua costituisca di gran lunga la maggior parte dell'entropia dell'universo. È stato stimato che se tutta la materia del nostro universo evaporasse, l'entropia totale cambierebbe soltanto di una frazione dell'uno per cento. Tutto questo ha notevoli conseguenze termodinamiche. Abbiamo visto che la formulazione classica del secondo principio prevede nel nostro futuro la morte termica (massima entropia); ma comprendiamo ora che all'inizio del nostro universo venne prodotta cosí tanta entropia che possiamo dire che la morte termica sta in effetti alle nostre spalle. Dobbiamo dunque analizzare piú da vicino i processi irreversibili dell'inizio dell'universo.

2. *Il modello cosmologico standard.*

Come si è detto, la moderna cosmologia è basata sulle equazioni di Einstein. Essa assume inoltre che l'universo sia uniforme e isotropo (questo è il cosiddetto principio cosmologico). Dato questo limite, le equazioni di Einstein prendono una forma molto semplice. Scriviamo queste equazioni senza dimostrazione per il caso piú semplice corrispondente alla pressione zero e allo spazio euclideo, in cui sono coinvol-

te solo due variabili. Una di queste, spesso chiamata «raggio» dell'universo R, determina i limiti che possiamo raggiungere attraverso l'osservazione astronomica; l'altra è la densità di energia ρ. Le equazioni di Einstein mettono in relazione ρ con il raggio R:

$$\text{(E.1)} \qquad \varkappa\rho = 3H^2$$

$$\text{(E.2)} \qquad \dot{\rho} = -3H\rho$$

dove $H = \dot{R}/R$ è la funzione di Hubble, e il punto indica la derivata rispetto al tempo cosmologico. La seconda equazione ha un significato particolarmente semplice: l'evoluzione dell'universo è adiabatica e reversibile; le equazioni di Einstein conservano l'entropia. Piú in generale, abbiamo tre variabili: il raggio dell'universo, la densità di energia ρ e la pressione p. Abbiamo dunque bisogno di un'equazione di stato $p(\rho)$. Dobbiamo tener conto anche della curvatura dello spazio.

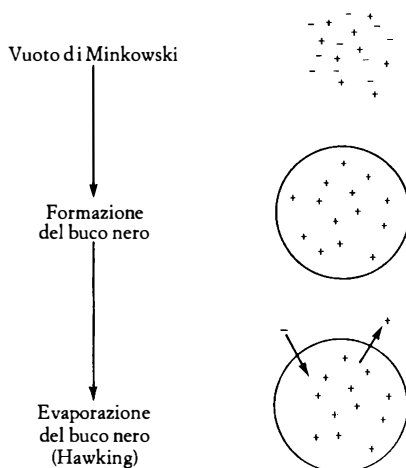
Per universi omogenei e isotropi, vi sono solo tre possibili casi, corrispondenti a curvatura negativa, nulla o positiva. Dànno origine rispettivamente allo spazio aperto (curvatura spaziale negativa), allo spazio euclideo (curvatura spaziale nulla), e allo spazio chiuso (curvatura spaziale negativa). Tutte queste soluzioni iniziano con un universo concentrato in un singolo punto. Questa conclusione è stata definita da John A. Wheeler «la piú grande crisi della fisica». Infatti, quale può essere il significato di una tale singolarità? Se ripercorriamo all'indietro la storia dell'universo, giungiamo veramente a un punto oltre il quale nessuna estrapolazione è sicura, un punto oltre il quale le leggi della fisica cessano di essere valide? Abbiamo veramente a che fare con una singolarità primordiale, o dobbiamo considerare l'inizio dell'universo come il risultato di un qualche tipo di instabilità, associata a un fenomeno del tipo transizione di fase?

3. *I buchi neri.*

Qualunque sia la risposta, i fisici concordano sul fatto che questi eventi accaddero durante la cosiddetta èra di Planck, determinata dalle tre costanti fondamentali: c (la velocità della luce), $\varkappa$ (la costante gravitazionale), e h (la costante di Planck), dalle quali si può derivare il tempo di Planck $\sim 10^{-44}$ s, un tempo veramente molto breve), la massa di Planck ($\sim 10^{-5}$ g, molto grande rispetto alla massa di un protone), e la lunghezza di Planck ($\sim 10^{-33}$ cm, che è molto piccola). Non possediamo

ancora una teoria fondamentale sull'èra di Planck. Questa dovrebbe implicare una teoria quantistica della gravitazione, che incontra notevoli difficoltà. Ma è interessante che qualche aspetto di questa teoria si possa ottenere senza risolvere i problemi fondamentali associati alla teoria quantistica dei campi. Un esempio è la teoria dei buchi neri di Stephen Hawking, che porta a previsioni degne di nota (figura E.1).

Si può pensare a un buco nero come a una «membrana» che racchiude una regione dello spazio-tempo che non permette alla luce di sfuggire. I confini dei buchi neri vengono chiamati «orizzonti degli eventi»: ogni qualvolta la luce supera questo orizzonte, vi rimane prigioniera (da cui il termine buco «nero»). Ma, come Stephen Hawking e altri hanno mostrato, i buchi neri non sono cosí «neri»: essi irradiano quando sono incastrati nel vuoto, cioè in uno stato dello spazio-tempo in cui la densità della materia e dell'energia si annullano. Inoltre, essi irraggiano come un corpo nero con temperatura T, che dipende solo dalla massa del corpo nero. L'origine di questa radiazione termica viene dalla loro interazione con il vuoto quanto-meccanico. Come abbiamo già detto nel capitolo V, § 10, il vuoto quanto-meccanico è il luogo delle fluttuazioni. Queste fluttuazioni portano alla formazione di coppie di particelle. Ma queste coppie sono coppie virtuali; esistono solo per un periodo molto breve. Una delle particelle ha energia positiva, l'altra ha energia negati-

Figura E.1

va. La radiazione di Hawking risulta dal fatto che un corpo nero assorbe particelle di energia negativa. Dunque, abbiamo un flusso di energia che fuoriesce dal buco nero, e il buco nero evapora. I buchi neri portano dunque a processi irreversibili.

4. *Il ruolo dell'irreversibilità.*

Veniamo ora al problema di vedere se l'origine del nostro universo sia una singolarità o un'instabilità. Un'idea che viene spesso presentata è che il nostro universo sia, come dicono gli economisti, un «free lunch» (pasto gratuito); ciò deriva dal fatto che tutta l'energia a disposizione sia alla fine presente sotto due forme fondamentali che si compensano l'una con l'altra: l'energia correlata alla massa (che è positiva) e l'energia correlata alla gravitazione (che è negativa). Cosí, dal punto di vista dell'energia, non c'è differenza fra il vuoto (l'universo di Minkowski) e l'universo materiale.

Abbiamo precedentemente discusso l'instabilità di Bénard. Anche in quel caso il prezzo da pagare non viene dato in termini di energia; in regime di convezione stazionaria, l'energia viene naturalmente conservata, e in questo senso abbiamo un «free lunch». È la produzione di entropia che fa la differenza. Viene dunque la tentazione di pensare che il prezzo da pagare per la nascita dell'universo sia la produzione di entropia. Per spiegare ciò, dobbiamo studiare piú attentamente l'impatto della creazione della materia sia sulla base dei due principî della termodinamica che sulla base delle equazioni di Einstein.

Consideriamo un volume V contenente N particelle. Per un sistema chiuso, N è costante. La conservazione dell'energia interna E viene dunque espressa dalla

$$dE = dQ - p\,dV \tag{E.3}$$

dove dQ è il calore ricevuto dal sistema durante un tempo dt. Possiamo riscrivere la relazione della equazione E.3 nella forma

$$d\left(\frac{\rho}{n}\right) = dq - pd\left(\frac{1}{n}\right) \qquad \text{dove} \qquad \rho = \frac{E}{V}, \quad n = \frac{N}{V}, \quad dq = \frac{dQ}{N} \tag{E.4}$$

Poiché la relazione dell'equazione E.4 è locale, essa rimane valida per sistemi aperti in cui N è dipendente dal tempo. In questo caso, l'equa-

zione E.3 viene modificata in modo da tenere conto della variazione del numero di particelle:

(E.5) $$d(\rho V) = dQ - p\,dV + \frac{h}{n} d(nV)$$

dove $h = \rho + p$ è l'entalpia per unità di volume. Per sistemi chiusi, le trasformazioni adiabatiche ($dQ = 0$) vengono definite dalla relazione

(E.6) $$d(\rho V) + p\,dV = 0$$

L'estensione ai sistemi aperti è data dall'equazione

(E.7) $$d(\rho V) + p\,dV - \frac{h}{n} d(nV) = 0$$

In questa trasformazione, il «calore» ricevuto dal sistema è unicamente dovuto al cambiamento del numero di particelle. Nella nostra prospettiva cosmologica, questo cambiamento è dovuto al trasferimento di energia dalla gravitazione alla materia. Dunque, *la creazione della materia agisce come una fonte di energia interna*. Ci occupiamo ora del secondo principio della termodinamica (cfr. capitolo II, § 5). Al solito, scomponiamo il cambiamento di entropia nel flusso di entropia d_eS e nella creazione di entropia d_iS:

(E.8) $$dS = d_eS + d_iS \qquad \text{con} \qquad d_iS \geqslant 0$$

Per valutare il flusso di entropia e la produzione di entropia, partiamo dal differenziale totale di entropia

(E.9) $$T\,d(sV) = d(\rho V) + p\,dV - \mu\,d(nV)$$

(E.10) $s = S/V$, μ è il potenziale chimico, $\mu n = h - Ts$, e $s \geqslant 0$.

Per sistemi chiusi e trasformazioni adiabatiche la relazione nell'equazione E.9 porta a

(E.11) $$dS = 0 \quad \text{e} \quad d_iS = 0$$

Consideriamo l'effetto della creazione della materia. Consideriamo sistemi omogenei e ci aspettiamo dunque di avere ancora $d_eS = 0$. In

contrapposizione, conveniamo che la creazione della materia contribuisca alla produzione di entropia. Abbiamo dunque:

(E.12) $$T\,d_iS = T\,dS = \frac{h}{n}d(nV) - \mu\,d(nV) = T\frac{s}{n}\,d(nV) \geqslant 0$$

Nel contesto cosmologico, questa disuguaglianza implica che lo spazio-tempo possa produrre materia, mentre il processo inverso è vietato dalla termodinamica. La relazione fra spazio-tempo e materia cessa di essere simmetrica, poiché la produzione di particelle, che avviene a spese dell'energia gravitazionale, risulta un processo irreversibile. L'interpretazione microscopica di questo processo in cui lo spazio-tempo viene convertito in materia verrà brevemente discussa alla fine di questa appendice.

Si può scrivere la relazione dell'equazione E.7 in varie forme equivalenti come:

(E.13) $$\dot{\rho} = \frac{h}{n}\dot{n}$$

(E.14) $$p = \frac{n\dot{\rho} - \rho\dot{n}}{\dot{n}}$$

È interessante notare che la creazione di energia $\dot{\rho}$ e la creazione di particelle determina la pressione p. Come esempio, notiamo che $\rho = Mn$ implica $p = 0$ e ancora, $\rho = \alpha T^4$, $n = bT^3$ implica $p = \rho/3$.

Per discutere le equazioni di Einstein per sistemi aperti, è utile trascrivere la relazione dell'equazione E.7 in una forma simile a quella dell'equazione E.6, ovvero

(E.15) $$d(\rho V) = -(p + p_c)\,dV$$

dove

(E.16) $$p_c = -\frac{h}{n}\frac{d(nV)}{dV} = -\frac{\rho + p}{n}\frac{d(nV)}{dV}$$

corrisponde a una pressione supplementare dovuta alla creazione di materia. A causa dell'equazione E.12, p_c è sempre negativa. Possiamo ora indicare il cambiamento delle equazioni di Einstein correlato alla creazione di materia. Quando c'è creazione di materia, dobbiamo introdur-

re la corrispondente pressione $\tilde{p}$ nelle equazioni di Einstein come la somma

$$\tilde{p} = p + p_c \tag{E.17}$$

della pressione usuale p piú la pressione p_c dovuta alla creazione di materia.

Otteniamo in questo modo una classe di nuove cosmologie, in cui le variabili sono R, r, p, e n. Consideriamo un semplice caso

$$\rho = Mn \tag{E.18}$$

che implica $p = 0$ per sistemi aperti. Abbiamo solo due variabili R e n. Una equazione è fornita dall'equazione E.1. Abbiamo bisogno di un'altra equazione, che colleghi la creazione di particelle con la funzione di Hubble H. Prendiamo per il volume V il valore

$$V \sim R^3 \tag{E.19}$$

L'equazione cinetica piú semplice che soddisfa la disuguaglianza nell'equazione E.12 è

$$\frac{1}{R^3}\frac{dnR^3}{dt} = \alpha H^2 \geqslant 0 \tag{E.20}$$

dove $\alpha \geqslant 0$. Per $\alpha = 0$ ricuperiamo l'usuale modello standard di descrizione, che include la singolarità del Big Bang. Tuttavia, per $\alpha \neq 0$, la situazione cambia drasticamente. Possiamo risolvere le equazioni E.1 e E.20, e mostrare che, qualunque siano i valori iniziali, il sistema tende allo *spazio di de Sitter*, caratterizzato da n e H costanti nel tempo. Ciò corrisponde a un'espansione esponenziale dell'universo, non appena $H = \dot{R}/R$. Questo risultato è del tutto indipendente dalla forma dell'equazione E.20. L'introduzione della creazione della materia rende dunque il Big Bang strutturalmente instabile, e la piú piccola creazione della materia porta a un universo che si espande esponenzialmente. A un livello piú microscopico, la creazione della materia dal vuoto può essere vista come un effetto dell'«instabilità» del vuoto rispetto alle sue fluttuazioni. Abbiamo visto che il vuoto quantistico è associato alla comparsa di particelle virtuali. Se un numero sufficientemente grande di queste particelle compare in un dato volume (come risultato di fluttuazioni), ciò può portare alla formazione temporanea di un buco nero, che

dovrebbe poi decadere, secondo il meccanismo di Hawking. In questo modo, le particelle virtuali vengono trasformate in particelle reali. Potremmo dunque intravedere una storia cosmologica in tre stadi: primo, l'instabilità del vuoto da cui emerge inevitabilmente un universo di de Sitter; poi, il processo dell'evaporazione del buco nero, che presumibilmente inibisce una ulteriore creazione; e infine, il passaggio dalla fase de Sitter alla fase dell'attuale nostro universo che si espande adiabaticamente, come descritto dal modello standard.

Una trattazione qualitativa di questa cosmologia introduce come parametro fondamentale la massa dei buchi neri m_{bh} che agiscono come intermediari fra le particelle virtuali e quelle reali. Per $m_{bh} \sim 100\ m_{Planck}$, si ottiene un perfetto accordo con il valore attuale delle quantità termodinamiche fondamentali del nostro universo, come il rapporto fra barioni e fotoni. Poiché la vita media dei buchi neri è molto breve, i due primi stadi corrispondono anche a un tempo molto breve, dell'ordine di 10^{-37} s, che è 10^7 volte piú grande del tempo di Planck.

È interessante notare che la formulazione quantistica di questo problema dà un ordine di grandezza simile per la massa dei buchi neri durante il processo di formazione del nostro universo. In termini del nostro modello, avremmo qualche tipo di transizione di fase, che separa la materia dalla gravitazione. Queste due componenti del nostro universo erano intimamente mescolate nel vuoto primordiale. La cosmologia moderna mescola ampiamente gli aspetti microscopici con quelli macroscopici. Non potremmo ottenere una descrizione degli stadi iniziali dell'universo senza la costante di Planck che regola il mondo microscopico. Questo esprime meravigliosamente ciò che viene spesso chiamato il carattere olistico dell'universo. Naturalmente, ciò di cui si discute qui non è la creazione dal nulla: il vuoto di Minkowski è già un mezzo ben definito, caratterizzato dai valori delle tre costanti fondamentali: c (la velocità della luce), $\varkappa$ (la costante gravitazionale), e h (la costante di Planck). Naturalmente si potrebbero porre delle domande piú ambiziose, come: da dove vengono queste costanti? E sono esse invarianti rispetto al tempo o stanno esse stesse cambiando? Non vi è allo stadio attuale nessuna guida teorica per rispondere a queste domande.

All'inizio di questa appendice, abbiamo citato la dualità del mondo newtoniano. Abbiamo enfatizzato il ruolo unificante della relatività generale di Einstein. Possiamo guardare alle equazioni di Einstein da un punto di vista analogo a quello di un ingegnere che sottolinea l'equivalenza fra calore e lavoro meccanico: le equazioni di Einstein esprimono un tipo di equivalenza fra lo spazio-tempo e la materia. Ma come nel ca-

so del calore e del lavoro, questa non è un'equivalenza piena. Lo spazio-tempo appare come una «forma nobile» dalla quale nasce la materia attraverso un processo irreversibile che implica una grande produzione di entropia. Uno dei temi ricorrenti di questo libro è che la produzione di entropia implica la produzione sia di ordine che di disordine, in accordo con la evidente dualità dell'universo.